

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

STIC-Biotech/ChemLib

77212

From: Davis, Minh-Tam
Sent: Friday, October 04, 2002 9:00 AM
To: STIC-Biotech/ChemLib
Subject: Search request for 09/631863

Please search in commercial database and in issued patent files:

SEQ ID NO:2

Thank you.

MINH TAM DAVIS

ART UNIT 1642, ROOM 8A01, MB 8E12

305-2008

priority
8/99

Point of Contact:
Toby Port
Technical Info. Specialist
CM1 6A04
703-308-3534

RECEIVED
OCT -4 2002
STIC

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 10/4
Date Completed: 10/4
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:

NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)

STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2002, 15:38:18 ; Search time 20.92 Seconds
(without alignments)
1841.865 Million cell updates/sec

Title: US-09-631-863A-2
Perfect score: 2139
Sequence: 1 MRNRKRVLTAKRRSGRGGD.....DNCPRKASKSSPAGNSPAPL 401

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	720.5	33.7	232	2 JE0163	myelin expression
2	240.5	11.2	853	2 S60178	gag polyprotein ho
3	190	8.9	457	2 T18347	gag protein homolo
4	159.5	7.5	639	2 S23569	gag polyprotein ho
5	148	6.9	1494	2 T13798	hypothetical prote
6	137	6.4	349	2 T18349	probable gag prote
7	135	6.3	537	1 F0MVRV	gag polyprotein -
8	134.5	6.3	555	2 T10349	structural protein
9	134	6.3	1529	2 A59310	unconventional myo
10	131.5	6.1	538	2 S70394	gag polyprotein -
11	131.5	6.1	1230	2 T18256	serine/threonine p
12	131.5	6.1	1230	2 T18259	hypothetical prote
13	130	6.1	334	2 T01815	hypothetical prote
14	130	6.1	537	1 F0MVRV	gag polyprotein -
15	129	6.0	1110	2 T19673	hypothetical prote
16	129	6.0	1188	2 T46608	zinc finger protei
17	125	5.8	915	2 T26695	hypothetical prote
18	125	5.8	1651	2 T14160	transmembrane rece
19	123.5	5.8	1182	2 T30189	myelin transcripti
20	123.5	5.8	1585	2 T31611	hypothetical prote
21	122.5	5.7	536	1 F0MVRV	gag polyprotein -
22	122.5	5.7	538	2 S35474	gag polyprotein -
23	122.5	5.7	601	2 S33377	gag polyprotein -
24	122.5	5.7	1316	2 T00381	pe3 protein - huma
25	121.5	5.7	1187	2 T46637	KIA0633 protein -
26	121.5	5.7	1585	2 T18274	transcription fact
27	121	5.7	1612	2 T30805	1-phosphatidylinos
28	120	5.6	745	2 D96829	duftl protein - mo
29	120	5.6	747	2 S71478	homeobox protein (
					homeotic protein A

30	120	5.6	992	2 T46337	hypothetical prote
31	120	5.6	1905	2 T18267	multidrug resistan
32	120	5.6	1937	2 T38055	myosin heavy chain
33	119.5	5.6	1870	2 S37671	MHC class III hist
34	119.5	5.6	1872	2 S36152	MHC class III hist
35	119.5	5.6	2854	2 T14156	MHC class III hist
36	119	5.6	320	2 AE2842	kinesin-related pr
37	119	5.6	359	2 F97619	conserved hypothet
38	119	5.6	465	2 A02986	hypothetical prote
39	118.5	5.5	428	1 I36930	myosin alpha heavy
40	118.5	5.5	7962	2 I38346	involucrin - white
41	118	5.5	2175	1 S03170	elastic titin - hu
42	117	5.5	972	2 T49773	homeotic protein c
43	116.5	5.4	249	2 A37280	related to actin-1
44	116.5	5.4	708	2 I83196	C/EBP-related prot
45	116.5	5.4	2142	2 B35098	NEDD-4 ORF - mouse
					MHC class III hist

ALIGNMENTS

RESULT 1
JE0163
myelin expression factor-3 - mouse
N/Alternate names: Myef-3
C/Species: Mus musculus (house mouse)
C/Date: 03-Jul-1998 #sequence_revision 10-Jul-1998 #text_change 07-May-1999
C/Accession: JE0163
R/Stephenski, A.; Krynska, B.; Treilakova, A.; Haas, S.; Knaill, K.; Ambul, S.
Biochem. Biophys. Res. Commun. 243, 295-301, 1998
A/Title: Myef-3, a developmentally controlled brain-derived nuclear protein which spe
A/Reference number: JE0163; MUID:98139908
A/Accession: JE0163
A/Molecule type: mRNA
A/Residues: 1-232 <STE>
A/Experimental source: brain
C/Keywords: phosphoprotein
F/40-60/Domain: transmembrane #status predicted <TM>
F/31,175/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status predi
F/31,142,185,201/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #stat

Query Match 33.7% Score 720.5 DB 2 Length 232:

Best Local Similarity 61.6% Pred. No. 1.6e-38;
Matches 149; Conservative 24; Mismatches 52; Indels 17; Gaps 5;

QY	154	CPEDLPKFDGNDPMLAPFMAQCQIFMEKSTRDPSVDVRCFVTSMNTGRA--ARMASA 211
DB	2	CTEDLPKFDGNDPMLGPRMYCCQIFMEKSTRDPSVDVRCFVTSMNLGRAPLGCOQA 61
QY	212	KL-----ERSHYLMHNYPAFMEMKHFVEDPQREYAKRRIRRLROGMSVIDYSNAFQMI 267
DB	62	KMYLDAOLHCLYDGAFAVLV-----PSESVSVQTDQTSAPGPGVVDYSNAFQMI 114
QY	268	ADDDLMNPEPALIDYHGEISDHOESLHLEFAKSLALIGCCHIERIRLAAAAARPR 327
DB	115	ADDDLMNPEPALIDYHGEISDHOESLHLEFAKSLALIGCCHIERIRLAAAAARPR 173
QY	338	SPRALVLPHTASHQVDPTEPVGARMLTQEEKERRKMLICVCTGCHYADNCPAK 387
DB	174	PSPRALVMP---PNSQTPTEPVGARMLTQEEKERRKMLICVCTGCHYADNCPAK 230
QY	388	AS 389
DB	231	AS 232

RESULT 2
S60178
gag polyprotein homolog - fungus (Fusarium oxysporum) retrotransposon skippy
C/Species: Fusarium oxysporum
C/Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 21-Jul-2000
C/Accession: S60178

R:Anaya, N.; Roncero, M.I.G.
Mol. Gen. Genet. 249, 637-647, 1995
A:Title: Skippy, a retrotransposon from the fungal plant pathogen *Fusarium oxysporum*
A:Reference number: S60178; MUID:96132549
A:Accession: S60178
A:Molecule type: DNA
A:Residues: 1-853 <ANNA>
A:Cross-references: EMBL:L34658; NID:G510695; PIDN:AAA88790.1; PID:G510696
A:Experimental source: retrotransposon skippy
C:Genetics:
A:Mobile element: retrotransposon skippy
C:Keywords: polyprotein

Query Match	11.28;	Score 240.5;	DB 2,	Length 853;
Best Local Similarity	21.28;	Pred. No. 1.2e-07;		
Matches 85; Conservative	65;	Mismatches 157;	Indels 93;	Gaps 14;

QY 33 GRSPTPTVTLGDCPPPPPPPPNNNNNNNSKHTGHSACVYNNMTERRDEL---SEEL 89
Db 19 GHQPAAP-----ANPAAPVNRPPPTDQNMDDADADSSQSSD--DSEVERLREQLGNVTNN 69

```

QY      90  NNLEKVV-----MKQSEENNNTLQSOVKLTRENTTLR--EQVEPTPEDEDDDIELRG 139
      |:::|      |::|::|::|      |      |:::|
Db      70  NEMKMLEEFTALQHOONOSNNNTQOEYNNLASAANGRDGCEYVAKPSP----- 118

```

```

QY 140 AAAAAAPPPIEECEBDLPKFDGNDMLAPFMAQCQTFMEKSTIRDFSVDRVRCVTS 199
    | | | | | : : : | : : :
Db 119 -----PEYEDGTSPKLPFLTOSRAFITYPNOFRNDSAKVMYVAG 159

```

```

OY      200  MMTPGAARMA SAKLESHYLMNYY-----PAFMME-----MKHVFEEDPQ 238
      :  ||:|  :||:|  |::|  :||:|  :||:|  :||:|  :||:|  :||:|  :||:|
Db      160  RLTKTAQWEP-----IMNDYMTNPYKIDORTALLEGNGRHEMEFALKMAEGTID 212

```

Qy 239 RREVAKRIRRLRGMSVIDISNAFOWIAQDLDMNEPALIDYHGLSHIQEELSHLE 298
:
: :|::|: |
213 EKGOAEPRITATKO-NGSASTICVEFFLOASKRDWDODVIMSEEFERALKEDVOOFTMEKN 271

QY 299 VAKSIALIGCIIERR---LARAAMARKPRSPRALVLPRIASHHQVDP-EPVGG 352

```

Qy 353 ARMRLTQEEKERRKLNLCLYCGGCHYADNC--PAKASK 390Q
      : : : : | | | | : :
Db 332 MAlGMKRRDKS----VTCYNGCKGKHYEACKNPVATNQ 367T

```

RESULT 3
T18347

C:Accession: T18347
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000
C:Species: Magnaporthe grisea (rice blast fungus)
C:Project: Magnaporthe grisea (rice blast fungus)
C:Project: Magnaporthe grisea (rice blast fungus)

Rafimanzana, M.B., Iosif, L., Milla, N., Moiré, Gen. Genet. 251, 665-674, 1996
A1: title: MAGGY, a retrotransposon in the genome of the rice blast fungus Magnaporthe grisea
A1: Reference number: Z18882; MUID:96335141

A;Accession: U1834/
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-457 <FAP>

A:Cross-references: EMBL:U5053; NID:g5222300; PID:g5222301; PIDN:AAA33419.1
C:Genetics:
A:Gene: gag
A:Webfile element: magnetoflo gungr retrotransposon

Query Match	Score	DB 2	Length
8.98	190	2	457

	77	CONDUCTANCE	29	MECHANICS	1007	ANALYSIS	687	SPE
OY	78	TERRADELSEINLR-----EKVMOSEENNNOISOVOKITEENTTL	120					
		: : :		: :		:		

Db	70	TDRKDAQAADILLMAMEKASATISRDAARELEKTESELEKVERTVLLPVLHITTPA	129
Oy	121	REOVEPTPEDEDDITELRGAAAAAARPPLE-----EBCREDLPREKDGNDMLAP	172
Db	130	NTPEP-----LMTVPMGTTPPASHENPASARLSSELP--PDKFTGASDRLRF	177
Oy	173	MAOCQIFEMKSTRFDSVDVAVCVTSMTGSAARMAASAKLESHYLMHYPARFEMKH	232
Db	178	ATQIRGKMTSKKDRPNEPSRLIYIAGLSCKAVNLLPKWVGSTPQFGDYTDLOYEE	237
Oy	233	VFEEDQREVEAKKIRTLRGMSGVIDYNAFQMTADDLMNEFALLDOYHEGLSHIOE	292
Db	238	AFGDDHVAONAKLYAKORNVDPAEYLSFEQSLSEGEPRFEDALPRLPFGISELOD	297
Oy	293	ELSH-----LEVAKSLSLA-----OCIIERRLAAARAAARPRSP-----PRA	332
Db	298	MILNHPASRQYHETTRLHLOSLDKRYRHOQYKKNQRTTPRAAARPARAARTQDIPRA	357
Oy	333	LVLPHIASHNOVDPERGVGAKMLTDEKKRRKKMLICLTGGIGHYADNCP	385
Db	358	AKPRAA-----ELPLNDP-----MLDSORRNRRNNMLCTYCSQGEHVAKCP	401

RESULT	4
S23569	

C:Accession: S23569; S19849
C:Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change 21-Jul-2000
C:Species: Cladosporium fulvum
gag polyprotein homolog

A. Reference number: S23569; MUID:92318885

A:Accession: 52309
A:Molecule type: DNA
A:Residues: 1-639 <MCH>
A:Cross-references: EMBL:211866; NID:q2562; PIDD:CAA77890.1; PID:q2563

A:Note: the authors translated the codon ACC for residue 55 as Ala, but for residue 56 as Thr.
C:Genetics:
A:Mobile element: retrotransposon Cfr-1
C:Superfamily: Gladosorinus fulvus probable and polyprotein

Query Match	Score	DB 2	Length
7.58	159.5	639	

Oy	159	PERFDGNDPLAPMAQOIFEMESTRDESVDRKVCFTVTSMTGTGRAARMAASAKLERSHY	218
Matches	65;	Conservative	37;
Mismatches	114;	Indels	55;
Gaps			9

Db 19 PEEFGDVRVKEPTWWSQMDMYEENS---MTENLKPIFAITTELRGRAQHWYKPELR--Y 73

QY 219 LMHN-----YPAFMEMKHFVEDPQRREYAKRKIRRLRQSGSVIDYSNAFOMI 267

Db 74 LDSNEDNADGVFKSYNHLKHAHKSVMFVSNELATAVRVIQHLTQ - KTSTAEYAAKFOEY 1332

Dd 133 AQLTMDDEALQVMYRRGLKEHKVDLMDGRKIDSLGDLVQTIDLDDKLTERAMERY 192

Ov 325 -----KRSPPRALVLPIASHHGVDPTEPYGGARMRLT 356

Db	193	DSKVS	KAGYTP	PGYD	NNRGR	GFND	NYNKR	PKDP	PIYGPQ	----	MELD	YTEK	----	GR	KTNS	245
Qv	359	QEFFE	----	RRRK	INLCT	YC	GTG	GHV	ADNC	PAK	387					

Db 246 KGNRRPSSRETRTCYGCGKPGHIARDCK 276

T13798
hypothetical protein - fruit fly (*Drosophila melanogaster*) retrotransposon mdg3
C:Species: *Drosophila melanogaster*

C:Date: 13-Nov-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13798
R:Avetisov, S.
Submitted to the EMBL Data Library, February 1996
A:Reference number: Z17761
A:Accession: T13798
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1494 <A>
A:Cross-references: EMBL:X95908; NID:e990667; PID:e223896; PIDN:CAA65152.1
C:Genetics:
A:Cross-references: FlyBase:Fgn0002698
A:Module element: retrotransposon mdg3

Query Match	6.98;	Score 148;	DB 2;	Length 1494;
Best Local Similarity	21.38;	Pred. No. 0.16;		
Matches 89;	Conservative 53;	Mismatches 155;	Indels 120;	Gaps 22

```

Oy 39 PTVLTGPDCCPP-PPPPPPNNN-----NNNSKHTGHSACV 74
      || ||| | ||: || | || |
Db 41 PTAVRG-DCPEHPQKNAPRGNDLSSLDFOCEINTDHSVAMNRKESTETG--SERE 97

```

```

Oy      75 PMTF--RRDELSEE---INLRKVKKOSEF-----NNLQSOYOK-----LTEEN 117
      || | : | | : | | : | | : | | : | | : | | : | | : | |
Db      98 TMMELQQLRRELAEKAMLNGTRSSLOFQEQQOPEQSKATVSSVIQTQPTQAGATKEN 157

```

```

Oy 118 TTTLREQVEPPPEDEDDDIELRGAAAAAAPPRIEECPEDLPKFKDGNPDLAPFMAQCQ 177
    || - - - - - :: :: || :
Db 158 TTTFHSPPQSRNRAESQRFVDAALA-----KETITDYDKT-----CARAW 199

```

[illegible]

```
QY      230  ---MKHVF-EDPQRREYAKRKIRRLRGOMGSYIDYSNAFQMAIQDLDWNEPALIDQYHEG  285
      : | : : : : | : : : : : : : : : : : : : : : : : : : : : : : :
Db      248  LDOLSLTFGEOSKAEIRRKFPESRKWKTEENFCSYDEKAKLSNGINIDDELDMIEG  307
```

```

QY      286 LSDHQEELSHLEVARSKSLALIGQCIIHERRLARA-AAARKP--RSPPRALVPLHTASHH 3422
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      308 IP-LONFRQARI-----QCFSPSEMLRAFSNIRLPAREEP----- 3444

```

```

QY 343 QVDPTEPVGARMLTQEE-----KERRRLNLCLYCGTGGHYADNCAPAKASKSS 392
    |::| |:::| | | | | | | |
db 345 -VOPDDYKDAIRCANCNSRGHKAIDCKKPKRESGSCYACGOLGHIVAOCPIRKSYSS 400

```

RESULT 6
T18349

probable gag protein - rice blast fungus gypsy retroelement
C:Species: Magnaporthe grisea (rice blast fungus)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C:Accession: T18349

R: Dobinson, K. F.
submitted to the EMBL Data Library, September 1994
A: Description: Sequence of the grh retroelement.
A: Reference number: 719882

A:Accession: T18349
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:PostAccess: 1-740-2005

A:cross-references: EMBL:M77661; NID:g538065; PID:g538066; PIDN:AAA21441.1
C:Genetics:
A:mobile element: gypsy retroelement

Query Match 6.4%; Score 137; DB 2; Length 349;

Matches 80; Conservative 41; Mismatches 124; Indels 98; Gaps

QY 105 NLOSQVQKLTENTTLREQVEPTPEDEDDITELRGAAAAAPP--PIIEECPEDLPEKF 162Z

```

Db      27  DLOGHVALLQIGAPTVPAIAE-----ALQATLALPRKKLRD-----PPLY 66
QY      163  DGNPDMLAPFMAOQIEMEKSTFSDVDRVRCVFTSMGTGRAAMAS-----AKLERSH 217
        67  DGVF--ASFAMRCAMEYKLRDAD-----PLGDHRDQYEXIWMGLETSVQKVRSY 116
Db
QY      218  YLM-----HNYPAFMEMKVIYEDPQOREYAKKKIRLRQMG-SVIDYSNAQ---MI 267
        117  YEVEGRDGAVERYTDFLDLETTYDDPHKRAQALAELETLKKMPQSAQFAIPIERTLAT 176
Db
QY      268  AODDWMNERPALIDQYHEGSLHIDOE-----LSHLFEAKSLSALIGCIIH 312
        177  AGGLMADEVTTNELRFVSPRIEACVGRMGCGTYLIGAVAIIRVOADLEA-----IE 233
Db
QY      313  IERRLA--RAAARKPRSPRALYLP--HIAHQVDPTEPVGGARR-----356
        232  LDRREGPHRAGAAVAPR-PRKDEDTPTMTGVAA--GSRPNGARGRRRGPOTPSDTN 286
Db
QY      357  -----LTOEKEKRRRLNCLVCGTSGTHADNCPAKAS 389
        287  RRDTPRAQWVPWSDEYQRRRTGCLRGSGNHVADCTYAAA 329

```

RESULT 7
FOMRV

N:contains: core protein p15; core shell protein p30; inner coat protein p30
C:Species: radiation murine leukemia virus
C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 16-Jul-1999

A: Title: Nucleotide sequence of a radiation leukemia virus genome.
Virology 158, 88-102, 1987

A:ReferenceNumber: R9402, MOL.D/201000
A:Accession: A26183
A:Molecule type: DNA
A:Residues: 1-537 <MER>

C:Superfamily: mammalian retrovirus gag polyprotein I
A:Gene: gag
C:Genetics:
A:Cross References: GB:K03503; GB:M08449; MID:g352032; PIDN:AAA40310.1; P

F1:1-129/Product: core protein p15 #status predicted <p15>
 F1:130-214/Product: inner coat protein p12 #status predicted <p12>
 F1:215-477/Product: core shell protein p30 #status predicted <p30>

Query Match 6 3% Score 135 DB 1 Length 537
F4/8-53/Product: nucleoprotein p10 #status predicted <p10>

31 BCTHUBS-----FATGCSDB-----TTPMYTICDNCB-----DDBDDDDNNNNN 61
Best Local Similarity 19.1%; Pred. No. 0.32;
Matches 97; Conservative 58; Mismatches 173; Indels 180; Gaps

```
Db      |||   |||   | : : | |||
        71 RGRHHPDQVPYIVTWEAIAIEBPSPWVKPFVSPKLSLPTAPILPSPGSPTPPP----- 124
```

```

      :| | :|      :: : | ::| : ||:
Db 125 -----RSALYPALTP-----SIKRPSPQVLSDNGLIDLETDEPPYG 165

```

Db 166 EQGPPSPDGGDDREAEATYTSETAPSPMYVSR LRGKRDPPAADSTTSRAFP LRLGNGNQLQ 225

```

Db      226  YMFSSSDLYNWKNNPFSFEDPGKLTALIESVLTTHQPTWDDCOQ-----L 272

```

273 LGITLTGEEKQVLLKARKAVGNDGKPTQLPNEVNSAPFLERPMDYTTPEGRNHLVLY 332

```

OY 277 -----ALIDQYHEGL-----SDHIQE--- 292
Db 333 ROLLLAGLQAGSPNTLAKVGTGCGPNESPFAFLERLKEAFRRYTPYDEHGOETSV 392
OY 293 -----ELSHLE--VAKSISALIGOCIH-----E 314
Db 393 SMSFIWQSPADIGKLEKLELDKSKTLRDLYVREAEKIFNKRETPEREEREREETEEENE 452
OY 315 RRLAARAAAKRRSPRR-----ALVLPRIASHHQVDPEPYGAMRLTOEKEKERRRLNL 370
Db 453 RRRREDEQRERKRRRRRREKREMSKILATVVGQRODRO--GGEKKR--POLDKO----- 502
OY 371 CLYCGTGHYADNCPAKASKSSPAGNSP 398
Db 503 CAYCKEKGHWAKDCPKK--PRGPRGPRP 528

```

RESULT 8

T30349

structural protein p78-81 - Lymantria dispar nuclear polyhedrosis virus

C:Species: Lymantria dispar nuclear polyhedrosis virus, LdMNPV

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T30349

R:Kuzio, J.; Pearson, M.N.; Harwood, S.H.; Funk, C.J.; Evans, J.T.; Slavicek, J.M.; Rohl

Virolgy 253, 17-34, 1999

A:Title: Sequence and analysis of the genome of a baculovirus pathogenic for Lymantria d

A:Reference number: 220836; MUID:99124785

A:Accession: T30349

A>Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-555 <KUZ>

A:Cross-references: EMBL:AF081810; PIDN:AAC70187.1

Query Match 6.3%; Score 134.5; DB 2; Length 555;

Best Local Similarity 20.7%; Pred. No. 0.35; Mismatches 129; Indels 107; Gaps 16;

Matches 75; Conservative 51; Mismatches 129; Indels 107; Gaps 16;

```

OY 21 PGLPHRSEATAGRSP-----TPVTLGPDPPPPPPPPPPNNNNNNNNNNKHTGKSACV 74
Db 214 PASPPARQESPISGSAPEPIRQETPTGLFAAPPPPPPPPP-----PPPPLOQKSSAV 267
OY 75 -----PNMTERRDELSEIINLREKVKQSEBNNNLOSQVOK-----LTBEHT 118
Db 268 PPPPPPLPPPGAVDDDFEGEGEV---RPKPAERAPTDALFEIRRGVOLKPRATERAP 324
OY 119 T-----LRE--OYEPPEDED-----DDELRGAAAAAAPP- 148
Db 325 TYPDDALFAETROGVKLPKPAERADDEPPKSSRAPLLLEINRDKIKLKVAPRATERPA 384
OY 149 -----PIE-----EECPEDLPKFDGNDPLAPFMAOCCIEMEKSTRDFSV----- 190
Db 385 SATNNPMLQMLNKRLESMKSSAASESDAN-----YSSMSDMEEDSLRDA 431
OY 191 -RVVVCVTSMTGR-----AARWASAKLESHYLMNYPFMEMKHVFEDPQREYAK 244
Db 432 LRIKIALIGPRLSESEKRIAKRLAGAKLSAEKTLDELQARAIEPNPLISPYQLTAP 491
OY 245 RKIKRLQMGSVTD-----YSNAFMI--AODLWNNPEPLIDQYHEGSDHIOEELSH 296
Db 492 LYLHDLKLFSASVLDLFRNGAYETALEKLEALQVLDQAVSL-ORMHDDISTYVKLOKR 550
OY 297 LE 298
Db 551 LE 552

```

RESULT 9

A59310

unconventional myosin heavy chain - maize

N:Alternate names: MYO1

C:Species: Zea mays (maize)

C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 08-Sep-2000

C:Accession: A59310

R:Li, L.; Pesacreta, T.C.

submitted to Genbank, May 1999

A:Reference number: A59310

A:Accession: A59310

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-1529 <LID>

A:Cross-references: GB:AF104924; NID:g4733890; PIDN:AAD17931.2; PID:g4733891

C:Genetics:

A:Gene: MYO1

C:Superfamily: myosin MYO2; myosin motor domain homology

F:65-719/Domain: myosin motor domain homology <MMO>

Query Match 6.3%; Score 134; DB 2; Length 1529;

Best Local Similarity 22.3%; Pred. No. 1.2; 128; Indels 72; Gaps 14;

Matches 73; Conservative 55; Mismatches 128; Indels 72; Gaps 14;

```

OY 78 TERRDELSEIINLREKVKQSEBNNNLOSQVOKLIENTTLREQ---VEP----- 136
Db 105 SELNEELIKFESAEEKIQLOETVRLBEKATNMESEKVLROQAVLSPSKSLAY 1064
OY 127 -----TPBEDDDIELGAAAAAPPPI-----EECPEDLPKFDGNDPL 169
Db 1065 KSPFQKTPENG--ALNGEVKSSPDITPILPNPKLEAEKPKQSLNEKQOENDL 1121
OY 170 APFMAO-----COIF-MEKSTRDESVDRVVCFTSMGTGRAAMWASAKLERS 216
Db 1122 IKCVSQDLGFSSGKPIACILYRCLLMWRSEFVEYRTGV-FDRIITQIGSLAESDNDK 1180
OY 217 HYLHNPYPAFMEMKHVE-----DPQREYAKRRIRRLQGM-GSVIDYNAF--Q 265
Db 1181 AYLWLSNSTLLILLQRLTKTTGAGFTPPRRSSAASFGVFSGMRSPSAGAFMGSR 1240
OY 266 MIA-----QDLWNERPL-----IDQYHEGSDHIOEELSHLEWAKSLIGOCIHIER 315
Db 1241 LIGLGLDROVEAKYPALRKOQLTAFLEKTYGMIRNLK-----KEISPLGLGICIOAP- 1294
OY 316 RLAARAAARRSPRALVLPRIASHHQ 343
Db 1295 RTRASLILKGRSQANLAAQOTLIAHWQ 1322

```

RESULT 10

S70394

gag polyprotein - Friend murine leukemia virus (strain FB29)

N:Contents: core protein p15; core shell protein p30; inner coat protein p12; nucleop

C:Species: Friend murine leukemia virus

A:Variety: Strain FB29

C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Jun-2000

C:Accession: S70394

R:Perryman, S.; Nishio, J.; Chesebro, B.

Nucleic Acids Res. 19, 6950, 1991

A:Title: Complete nucleotide sequence of Friend murine leukemia virus, strain FB29.

A:Reference number: S70393; MUID:92107687

A:Accession: S70394

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: genomic RNA

A:Residues: 1-538 <PPR>

A:Cross-references: EMBL:Z11128; NID:g61547; PIDN:CAA77478.1; PID:g2654364

A:Experimental source: strain FB29

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991

C:Genetics:

A:Gene: gag

C:Superfamily: mammalian retrovirus gag polyprotein I

C:Keywords: core protein; inner coat protein; nucleoprotein; polyprotein

Query Match

6.1%; Score 131.5; DB 2; Length 538;

Best Local Similarity 18.8%; Pred. No. 0.53;

Matches 95; Conservative 53; Mismatches 185; Indels 171; Gaps 17;


```

21 POLAHR-----EATGRSP-----PRTVTLLGPDCCPPPPPPNNNNNNNS 64
   ||| |
   ||| |
Db 71 PGRPHRPPQVPIYVWELAVDPWPVVRFPVHKRPLSPAPSLPREPLS----- 123
   ||| |
   ||| |
Oy 65 KHTGKACVPMTERRRDELSEELNNIREKVKOSEENNLSQOVOKLTENTTLREOV 124
   ||| |
   ||| |
Db 124 -TPPOSSLYPALT-----SPLNTKPRQVLPDGS-----GPLIDLTEDDPPEYRDPG 169
   ||| |
   ||| |
Oy 125 EPTPEDEDDDELRCGAAAAAP-----PPTIEECPEDLPKFCQNP-MLAP 171
   ||| |
   ||| |
Db 170 PPSPDGNGDSGVAPTEGAPDPSVWVSRGRKEPPVADSTTSQAFLRLGNGQYQWP 229
   ||| |
   ||| |
Oy 172 FMAOQIEMEKSTRDPSVDVRYV-CFVTSMTGARAARMAKLESHYLMHNYPAFMM 230
   ||| |
   ||| |
Db 230 FSSSLVLMKNNNPFSEDPRAKTLTIESVLLTHQPTDDCO-----LGCT 276
   ||| |
   ||| |
Oy 231 KHVEFDPQREVAKRRIRLRQMGSVI-DYNAFQMLAODLDN----- 274
   ||| |
   ||| |
Db 277 LLTGEKORVLLKARKAVRGEDGRPTQLPNDINDAFPLERPMDYNTQGRNHLVHYRL 336
   ||| |
   ||| |
Oy 275 -----EPALIDYHEGL----- 286
   ||| |
   ||| |
Db 337 LLAGLQNGRSPNLAQVKGITQGNESPSAFLERLKAYRKYTPYDEPDGQETNVAMS 396
   ||| |
   ||| |
Oy 287 -----SDHIOELSHLE--VAKSLALIGOCIH-----ERRL 317
   ||| |
   ||| |
Db 337 FIMOSAPITGKLELEDKSTGLDLYREAKITNKRETPEREERIRRETEEKERRR 456
   ||| |
   ||| |
Oy 318 ARAAARPRSPRALVLRPHLASHHOVDPTPEVGARMLTOEKEERRKL---MLCLE 374
   ||| |
   ||| |
Db 457 AEDVQREKERDPRR-----HREMSKLATVVSQGRQDQGERRRRQLDHQQAYC 507
   ||| |
   ||| |
Oy 375 GTGGHYALNCRAKASKSPAGNSP 398
   ||| |
   ||| |
Db 508 KEKGWARDCPK--PRGPRGPR 529
   ||| |
   ||| |

RESULT 11
T18256
Probable serine/threonine-specific protein kinase (EC 2.7.1.-) - yeast (Candida albicans)
C:Species: Candida albicans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18256
R:Leberer, E.; Harcus, D.; Broadbent, I.D.; Clark, K.L.; Dignard, D.; Ziegelbauer, K.; S
Proc. Natl. Acad. Sci. U.S.A. 93, 13217-13222, 1996
A:Title: Signal transduction through homologs of the Ste20p and Ste7p protein kinases ca
A:Reference number: 218843; MUID:97075145
A:Accession: T18256
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1230 <LEB>
A:Cross-references: EMBL:L47210; NID:92276410; PID:92286042; PIDN:AA865439.1
C:Keywords: phosphotransferase; serine/threonine-specific protein kinase

Query Match 6.1%; Score 131.5; DB 2; Length 1230;
Best Local Similarity 21.5%; Pred. No. 1.4;
Matches 64; Conservative 51; Mismatches 121; Indels 61; Gaps 13;

Oy 17 GGDGGLPHR-SEATACRSPTPTVTLGPDCCPPPPPPNNNNNNNSKTHGNSA--- 72
   ||| |
   ||| |
Db 770 GGNNALPKORINERKARARPP--PSAPAPRPVPPAPNALSQTSSELPQOQTAISQ 827
   ||| |
   ||| |
Oy 73 -----CVPNMTERRRDELSEELNNIREKVKOSEENNLSQOVOKLTENTTLREOYEP 126
   ||| |
   ||| |
Db 828 ALADVTAPNTIYEIOQT KYGEOQKLRKKARELEE-----IQRLREKNRQROQET 880
   ||| |
   ||| |
Oy 127 TPDEDDDIELRGAAAAAAPPRIEECPEDLPKFCQNP-MLAPFAQ----- 175
   ||| |
   ||| |
Db 881 GONNAD-----TASGGSNIAPVPVNNKR--PSGSGGRGAKQALALIAOKKREKKRN 933
   ||| |
   ||| |
Oy 176 COIFEMKST-----RDFSVDVAVCFVTSMGTGRAARMAKLESHYLMHNPAPFM 228
   ||| |
   ||| |
```

```

Db 934 IQIILAKLTICNPDPNELYDVKI-----GQASGVFLAHVDKSNIVAI 982
Oy 229 EKHVFEDPQREVAKRRIRLRQMGSVIDYNAFQMLAODLDNNEPALDQYHEG 285
   ||| |
   ||| |
Db 983 KQMLEQPKKELLINELVMKSSHPNIVNFIDSY-LLKQDL-W-----VIMEYMEG 1033
   ||| |
   ||| |

RESULT 12
T18259
serine/threonine protein kinase homolog - yeast (Candida albicans)
C:Species: Candida albicans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-May-2000
C:Accession: T18259
R:Kohler, J.R.; Fink, G.R.
Proc. Natl. Acad. Sci. U.S.A. 93, 13223-13228, 1996
A:Title: Candida albicans strains heterozygous and homozygous for mutations in mltoe
A:Reference number: 211118; MUID:97075146
A:Accession: T18259
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1230 <KOH>
A:Cross-references: EMBL:U73457; NID:91657953; PID:91737181; PIDN:AA838875.1
C:Genetics:
A:Note: CSTR20

Query Match 6.1%; Score 131.5; DB 2; Length 1230;
Best Local Similarity 21.5%; Pred. No. 1.4;
Matches 64; Conservative 51; Mismatches 121; Indels 61; Gaps 13;

Oy 17 GGDGGLPHR-SEATACRSPTPTVTLGPDCCPPPPPPNNNNNNNSKTHGNSA--- 72
   ||| |
   ||| |
Db 770 GGNNALPKORINERKARARPP--PSAPAPRPVPPAPNALSQTSSELPQOQTAISQ 827
   ||| |
   ||| |
Oy 73 -----CVPNMTERRRDELSEELNNIREKVKOSEENNLSQOVOKLTENTTLREOYEP 126
   ||| |
   ||| |
Db 828 ALADVTAPNTIYEIOQT KYGEOQKLRKKARELEE-----IQRLREKNRQROQET 880
   ||| |
   ||| |
Oy 127 TPDEDDDIELRGAAAAAAPPRIEECPEDLPKFCQNP-MLAPFAQ----- 175
   ||| |
   ||| |
Db 881 GONNAD-----TASGGSNIAPVPVNNKR--PSGSGGRGAKQALALIAOKKREKKRN 933
   ||| |
   ||| |
Oy 176 COIFEMKST-----RDFSVDVAVCFVTSMGTGRAARMAKLESHYLMHNPAPFM 228
   ||| |
   ||| |
Db 983 KQMLEQPKKELLINELVMKSSHPNIVNFIDSY-LLKQDL-W-----VIMEYMEG 1033
   ||| |
   ||| |

RESULT 13
T01815
hypothetical protein T27D20.19 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 24-Mar-1999
C:Accession: T01815
R:Edwards, J.; Wollam, C.; Dubdelde, C.
submitted to the EMBL Data Library, August 1998
A:Description: The sequence of A. thaliana T27D20.
A:Reference number: 214441
A:Accession: T01815
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-334 <EDW>
A:Cross-references: EMBL:AF076274; NID:93293583; PID:93377854
C:Genetics:
A:Map position: 4
A:Introns: 79/1; 107/2; 133/2
A:Note: T27D20.19
```

[illegible]

```

RESULT      14
FROMVIB     gag polyprotein - murine leukemia virus (strain BM5 ECO)
NContains:  core protein p15; core shell protein p30; inner coat protein p12; nucleoprotein
C:Species:   murine leukemia virus
C:date:      30-Sep-1992 #sequence.revision 30-Sep-1992 #text.change 16-Jul-1999
C:Accession: A40416
R:Chattopadhyay, S.K.; Sen Gupta, D.N.; Fredrickson, T.N.; Morse III, H.C.; Hartley, J.W.
J.Virol. 65, 4232-4241, 1991
A>Title:     Characteristics and contributions of defective, ecotropic, and mink cell focus
A:Reference number: A40416; MUID:91303677
A:Accession: A40416
A:Molecule type: DNA
A:Residues: 1-537 <CHAF>
A:Cross-references: GB:M64095; NID:g332014; PIDN:AAA46510.1; PID:g332015
C:Genetics:
C:Gene:      gag
C:Superfamily: mammalian retrovirus gag polyprotein I
C:Keywords:   core protein, inner coat protein, nucleoprotein; polyprotein
F:1-129/Product: core protein p15 #status predicted <CP1>
F:130-214/Product: inner coat protein p12 #status predicted <ICP>
F:215-477/Product: core shell protein p30 #status predicted <CSP>
F:478-537/Product: nucleoprotein p10 #status predicted <NP1>

Query Match          6.1%; Score 130; DB 1; Length 537;
Best Local Similarity 18.5%; Pred. No. 0.65;
Matches    93; Conservative 53; Mismatches 186; Indels 172; Gaps 17;

Oy  21 POLHHR-----EATGRSP-----TPVTGDDCP-----PPPPPNNNNN 61
       ||| | | | | | | | | | | | | | | | | | | | | |
Db  71 PGPCHGPQOVPIYWEIAYEPWPWKPEVSPKISLPTAPILPSGSTOPP----- 124
                                : : : : : : : : : : : : : : : :
Oy  62 NNSKHTGHKSACVPNMTRRRRDSEELINLRKKMKOSEENNLOSOKLTBEENTTLR 121
       ||| | | | | | | | | | | | | | | | | | | | | |
Db  125 -----RSALYPAFT-----PSIKPRSKQVYLSDDGCL----IDLTEDEPPRYG 165
                                : : : : : : : : : : : : : : : :
Oy  122 EQVEPTPEDDEDIDLEKAAAAAP-----PPPIEECCEDLPKFDGNPDM- 168
       ||| | | | | | | | | | | | | | | | | | | | | |
Db  166 EOGPSSPDGDDEAREATSTSELIPASPMWSRLRGKRDPADSTTSRAFPRLGGNGQLQ 225
                                : : : : : : : : : : : : : : : :
Oy  169 LAPFAAOQIEFMKSTROFSVDRAKV-CFVTSMTMGRAARAASAKLESHTLMHNYPAFM 227
       ||| | | | | | | | | | | | | | | | | | | | | |
Db  226 YWPFESSDLVWNKNPNPSFSEDPGKLIALIESVLITTHOPTWDDCO-----V-L 272

```

QY	228	MEMHVEDPORRVARKRIIRLROGGSVI--DYSAPQMIDDDWNEP-----	276
Db	273	LGTLLTBEKORVLLERAKVAGNDKRPOTLPMEVNSAFLEPPMDWYTTPEGRNHLVLY	332
QY	277	-----ALIDQYHEGLSDH-----	289
Db	333	ROLLLAGLONAGRSPTNLAKVGITGGPNESPASFLERLKEAVRYTTPYDPEDDGQETNV	392
QY	290	-----IOELSHLE--VAKSLALIGOCIH-----E	314
Db	393	SMSFTWOSAPALIGKRLERLEDLKSKTGLDYLREAEKIFPKRTPPERBERIRIRETEEKE	452
QY	315	RRLRAAAARRRPSPRALYVLPNASHHOVDPTPEYVGAGMRLLTQEEKERRRKLNLCIYC	374
Db	453	RRRAGDEOREKERDRRRROREMSKLLATVTVGQRQDRGGERRRRPQDKDQ-----CAVC	506
QY	375	GTGHHYADNCPAKAKSSSPAGNSP	398
Db	507	KKGHWADCEPK--PRGPRGPR	578

RESULT 15
 T19673
 hypothetical protein C33B4.3 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T19673
 R:Coles, L.
 submitted to the EMBL Data Library, February 1995
 A:Reference number: Z19160
 A:Accession: T19673
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1110 <WIL>
 A:Cross-references: EMBL:Z48367; PIDN:CAA88324.1; GSPDB:GNO0020; CESP:C33B4.3
 A:Experimental source: clone C33B4
 C:Genetics:
 A:Gene: CESP:C33B4.3
 A:Map position: 2
 A:Introns: 20/3; 110/3; 144/3; 341/1; 455/1; 546/1; 653/3; 934/2; 983/3; 1035/1

[illegible]

Search completed: October 4, 2002, 15:39:11
Job time: 53 sec

Fri Oct 4 15:49:00 2002

us-09-631-863a-2.rpr

Page 7

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2002, 15:38:18 ; Search time 31.76 Seconds
(without alignments)
1402.412 Million cell updates/sec

Title: US-09-631-863A-2
Perfect score: 2139
Sequence: 1 MRNRVLTAKTKRRSGRGQD.....DNCPAKSKSPAGNSAPPL 401

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: A.Geneseq_032802:*
2: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT:*
3: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT:*
4: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT:*
5: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT:*
6: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT:*
7: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT:*
8: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT:*
9: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT:*
10: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT:*
11: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT:*
12: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT:*
13: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT:*
14: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT:*
15: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT:*
16: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT:*
17: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT:*
18: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:*
19: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT:*
20: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:*
21: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT:*
22: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:*
23: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1802	84.2	342	21	AA842148 Human ORFX ORF1912
2	245.5	11.5	1607	22	ABG15099 Novel human diagno
3	191	8.9	502	22	ABG09602 Novel human diagno
4	159	7.4	146	20	AA26051 Human alpha helica
5	159	7.4	146	22	AA860492 Human cell cycle a
6	156	7.3	520	22	AAU15111 Protein encoded by
7	155	7.2	126	20	AA25512 Human mature alpha
8	149.5	7.0	110	20	AA26054 Human zalfaphal epit
9	149.5	7.0	168	21	AA26054 Human zalfaphal epit
10	149.5	7.0	184	21	AA26054 Human zalfaphal epit
11	139.5	6.5	142	22	ABG12205 Novel human diagno

ALIGNMENTS

RESULT	ID	AA842148 standard; Protein; 342 AA.	ALIGNMENTS	Human cell cycle a
1	AA842148			Human myomectrium t
2	AA842148			Candida albicans c
3	AA842148			Human neurofilamen
4	AA842148			Drosophila melanog
5	AA842148			Human shear stress
6	AA842148			Drosophila melanog
7	AA842148			Drosophila melanog
8	AA842148			Drosophila melanog
9	AA842148			C albicans apoptos
10	AA842148			Human p160 polyep
11	AA842148			Human p160 polyep
12	AA842148			Drosophila melanog
13	AA842148			Novel human diagno
14	AA842148			Novel human diagno
15	AA842148			Novel human diagno
16	AA842148			Novel human diagno
17	AA842148			Novel human diagno
18	AA842148			Novel human diagno
19	AA842148			Novel human diagno
20	AA842148			Novel human diagno
21	AA842148			Novel human diagno
22	AA842148			Novel human diagno
23	AA842148			Novel human diagno
24	AA842148			Novel human diagno
25	AA842148			Novel human diagno
26	AA842148			Novel human diagno
27	AA842148			Novel human diagno
28	AA842148			Novel human diagno
29	AA842148			Novel human diagno
30	AA842148			Novel human diagno
31	AA842148			Novel human diagno
32	AA842148			Novel human diagno
33	AA842148			Novel human diagno
34	AA842148			Novel human diagno
35	AA842148			Novel human diagno
36	AA842148			Novel human diagno
37	AA842148			Novel human diagno
38	AA842148			Novel human diagno
39	AA842148			Novel human diagno
40	AA842148			Novel human diagno
41	AA842148			Novel human diagno
42	AA842148			Novel human diagno
43	AA842148			Novel human diagno
44	AA842148			Novel human diagno
45	AA842148			Novel human diagno

Human ORFX ORF1912 polypeptide sequence SEQ ID NO:3824.

RESULT	ID	AA842148 standard; Protein; 342 AA.	ALIGNMENTS	Human cell cycle a
1	AA842148			Human myomectrium t
2	AA842148			Candida albicans c
3	AA842148			Human neurofilamen
4	AA842148			Drosophila melanog
5	AA842148			Human shear stress
6	AA842148			Drosophila melanog
7	AA842148			Drosophila melanog
8	AA842148			Drosophila melanog
9	AA842148			C albicans apoptos
10	AA842148			Human p160 polyep
11	AA842148			Human p160 polyep
12	AA842148			Drosophila melanog
13	AA842148			Novel human diagno
14	AA842148			Novel human diagno
15	AA842148			Novel human diagno
16	AA842148			Novel human diagno
17	AA842148			Novel human diagno
18	AA842148			Novel human diagno
19	AA842148			Novel human diagno
20	AA842148			Novel human diagno
21	AA842148			Novel human diagno
22	AA842148			Novel human diagno
23	AA842148			Novel human diagno
24	AA842148			Novel human diagno
25	AA842148			Novel human diagno
26	AA842148			Novel human diagno
27	AA842148			Novel human diagno
28	AA842148			Novel human diagno
29	AA842148			Novel human diagno
30	AA842148			Novel human diagno
31	AA842148			Novel human diagno
32	AA842148			Novel human diagno
33	AA842148			Novel human diagno
34	AA842148			Novel human diagno
35	AA842148			Novel human diagno
36	AA842148			Novel human diagno
37	AA842148			Novel human diagno
38	AA842148			Novel human diagno
39	AA842148			Novel human diagno
40	AA842148			Novel human diagno
41	AA842148			Novel human diagno
42	AA842148			Novel human diagno
43	AA842148			Novel human diagno
44	AA842148			Novel human diagno
45	AA842148			Novel human diagno

PR 02-APR-1999; 9905-0127636.
PR 05-APR-1999; 9905-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
XX
PA (CURA-) CURAGEN CORP.
PI
PI Shinkets RA, Leach M;
DR
DR WPI: 2000-602362/57.
DR N-PSDB; AAC76357.
XX
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 11: Page 2977-2978; 5507pp: English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antiproliferative; antiparkinsonian; nontropic; immunoprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotrophic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antineoplastic; antibacterial; antiviral; antifungal; antineumatic;
CC antihypertensive; antianemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypermetabolism, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SO Sequence 342 AA;
Query Match 84.2%; Score 1802; DB 21; Length 342;
Best Local Similarity 100.0%; Pred. No. 1.8e-137;
Matches 342; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 60 NNNNSKTHGHSACVPMNTERRRDELSEINNLREKVKOSENNNNOSQOKLLEEMFT 119
Db 1 nnnnskthghksacvpmnterridelseeinlrekykqseennlqsqvqkileentc 60
QY 120 LREQVEPTPEDEDDDELKGAATAAAPPPIEEBCPEDLPKFDGNPDLAPFMAQCQIF 179
Db 61 lreqveptpeededdldelrgataaaapppleeepelpkfdgnpdlapfmaqcqif 120
QY 180 MEKSTRDFSVDRVRCFTSMGTGAARWASAKLERSHYLMHNPAFMEKKHVPEDQR 239
Db 121 mekstrdfsvdrvrcftsmgtgaarwasaaklershylmhnpafmmekehvfedpqr 180
QY 240 REVAKRKTRRLROGMSVYIDSNAFOMIADLDNNEPALIIOYHEGLSDHIQELSHLEV 299
Db 181 revakrktrrlrogmsvdyidynaftaiaqldvnepalidqyheglshlev 240
QY 300 AKSLALIGOCIHIERRLARAAAARKPRSPRALVLPRIASHHOVPTPEVGAGARMLTQ 359
Db 241 akslalsligncihierlraaaaarkprspralvlpriashhovptpevgagarmrltq 300
QY 360 EEKERRRLNLCLYGTGTGHTADNCPAKASKSSPAGNSPAPL 401
Db 301 eekerrrlnlclcygtgtgthadncpakasksspagnspapl 342
RESULT 2
ABGI5099

ID ABGI5099 standard; Protein; 1607 AA.
XX
XX ABGI5099;
AC
AC
DT 18-FEB-2002 (first entry)
XX
XX
DE Novel human diagnostic protein #15090.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
XX WO200175067-A2.
PN
XX
PD 11-OCT-2001.
XX
XX
PF 30-MAR-2001; 2001MO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
XX Drmanac RT, Liu C, Tang YT;
PT
XX
DR WPI: 2001-639362/73.
DR N-PSDB; AAS79286.
XX
PS Claim 20: SEQ ID NO 45458; 103pp: English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABC00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcr-sequences.
XX
SO Sequence 1607 AA;
Query Match 11.5%; Score 245.5; DB 22; Length 1607;
Best Local Similarity 25.1%; Pred. No. 7.9e-11;
Matches 99; Conservative 54; Mismatches 149; Indels 93; Gaps 15;
QY 9 TKRRSG---RGQDPGLHPRSEATAGRSPT-----PTVTLCPPDCPPPPPP-- 54
Db 282 tteatsgsvvgeagpasgp-----aqekkeppspplameelpldl1-qdmeegssgprk 336
QY 55 ----PPNN-----NNNNSKTHGHSACVPMNTERRRDELSEINNLREK-----VMKQ 99
Db 337 elcdpndllqldleescnsh-----gagdpdpsgaadrmkeasvnpnsgrae 384

Query Match	7.4%	Score 159	DB 20	Length 146
Best Local Similarity	27.4%	Pred. No. 3.4e-05		
Matches 45	Conservative 21	Mismatches 68	Indels 30	Gaps 2
OY	83	DELSEINLRKRVKWKOSENNLOSOKLLEENTTLREOVETPRPEDDDDIELRGAA	142	
DB	3	deylvllhalmthralstlensqneqirlllyceaslirgvrp-----	46	
OY	143	AAAPPPIEECPEDLPEKFDGNPDLAPFMAQCOIFEMKSTRDFSVDRVRCFTSMNT	202	
DB	47	-----pscpvpfpetfngesarrlpelivqtasymlvnenrfondamkvaflslilt	97	
OY	203	GRARWASAKLERSHYLMNYPAFMEKHKHYF-----EDPQKRE	241	
DB	98	geaeewvvyiemdspilgdyrafidemkqctgwddeedddee	141	
RESULT	5			
AAB60492				
ID	AAB60492	standard	Protein: 146 AA.	
XX	AAB60492			
DT	24-APR-2001	(first entry)		
XX				
DE	Human cell cycle and proliferation protein CCYPR-40, SEQ ID NO:40.			
XX				
XX	Cell cycle and proliferation protein; CCYPR; human; agonist;			
KW	antagonist; gene therapy; detection; gene therapy;			
KW	transgenic animal disease model; immune disorder;			
KW	developmental disorder; cell signalling disorder;			
KW	cell proliferative disorder; cancer; tumour; anaemia; epilepsy;			
KW	arteriosclerosis; asthma; allergy; diabetes mellitus;			
KW	menstrual cycle disorder; bacterial infection.			
XX				
OS	Homo sapiens.			
XX				
PN	WO200107471-A2.			
XX				
PD	01-FEB-2001.			
XX				

PF	21-JUL-2000; 2000OWO-US19948.
XX	
PR	21-JUL-1999; 99US-01545075.
FR	08-SEP-1999; 99US-0153129.
PR	10-NOV-1999; 99US-0164647.
XX	
PA	(INCY-) INCYTE GENOMICS INC.
XX	
PI	Hillman JL, Ial P, Tang YT, Yue H, Au-Young J, Bandman O;
PI	Azimzai Y, Yang J, Lu DAM, Baughn MR, Patterson C, Shah P;
DR	N-PSDB; AAF59629.
XX	
PT	WPI; 2001-112727/12.
XX	
PL	Human cell cycle and proliferation proteins and polynucleotides are
PT	used to treat, diagnose and prevent immune, developmental and cell
PL	signalling disorders and cell proliferative disorders including cancer -
XX	
PS	Disclosure; Page 150; 205pp; English.
XX	
CC	Sequences AAB60453-AAB60506 represent 54 human cell cycle and
CC	proliferation proteins (CCYP), which are encoded by AAF59590-AAF59643,
CC	CCYP and agonists of CCYP are used to treat diseases or conditions
CC	associated with decreased expression of functional CCYP, while CCYP
CC	antagonists are used to treat diseases or conditions associated with
CC	overexpression of functional CCYP. Monoclonal or polyclonal antibodies
CC	to CCYP may be used in enzyme-linked immunosorbent assays (ELISA) or
CC	radioimmunoassays to detect CCYP. CCYP itself may be used to detect
CC	compounds e.g., antibodies, oligonucleotides and proteins (receptors)
CC	that specifically bind to CCYP, and in drug screening methods to
CC	identify compounds that modulate the activity of CCYP. CCYP
CC	nucleotides can be used to generate transgenic animal models of human
CC	disease, and can be used in gene therapy in target cells with genetic
CC	abnormalities with respect to the expression of CCYP for the
CC	treatment or prevention of a disorder associated with CCYP.
CC	Diseases which can be diagnosed, treated and prevented using CCYP
CC	proteins, nucleic acids, agonists or antagonists include immune,
CC	developmental and cell signalling disorders, and cell proliferative
CC	disorders including cancer, specific examples of these disorders
CC	include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,
CC	diabetes mellitus, disorders of the menstrual cycle and infections
CC	caused by bacteria.
XX	
SQ	Sequence 146 AA;
XX	
OY	Query Match 7.4%, Score 159; DB 22; Length 146;
XX	Best Local Similarity 27.4%, Pred. No. 3.4e-05;
XX	Matches 45; Conservative 21; Mismatches 68; Indels 30; Gaps 2;
Db	83 DELSEINNNLEKVKKOEENNNNLSQVOKLTLENTTLREQVEPTPEDEDDDIELRGAA 142
Oy	: : : : : : : : :
Db	3 delvlllhalmrhalsiensqmeqlrllvcasallrgvr----- 46
Oy	143 AAAPPPLEECEPEDLPKKFDGNPMDLAPFAAOCEFEKESTRDFSVDRVCFVTSMT 202
Db	: : : : : : : : :
Oy	47 -----pscpvfpetingessrlpelivqltasynlvnenctdamkavafisllt 97
Oy	203 GRAARASAKLERSHYLMHNTPAFMEEKHYF-----EDPORRE 241
Db	: : : : : : : : :
Oy	98 gaeewvvpyiemdspligdyrafidemkgcfgywddeddaee 141
XX	
RESULT	6
ID	AAU15111
AC	AAU15111 standard; Protein; 520 AA.
XX	
AC	AAU15111;
XX	
DT	04-DEC-2001 (first entry)
XX	
DE	Protein encoded by C. albicans essential gene CatpPL228W (CET1).


```

Db      2  ensqqlmeqilrllycerasillrvrp-----pscpvpifpet 36
OY      162 FDGPPDMLAPFMAOCQIFEMKSTRDFSVDRVRCFVTSNMTGRARASAKLERSHYLMH 221
Db      37  fngessrlpeflvtqasymlvnenfcdamkvaflislltgeeevvpylendspllg 96
OY      222 NYPAFMEMKHFV-----EDPORRE 241
Db      97  dyrafldemkqcfqgwdeddddee 121

RESULT      8
AAV26054
ID  AAV26054 standard; Protein; 110 AA.
XX
AC  AAV26054;
XX
DT  28-SEP-1999 (first entry)
XX
DE  Human zalphal epitope-bearing protein fragment 3.
XX
KW  Alpha helical protein-1; Fragile-X syndrome; cardiovascular system;
KW  connective tissue; abnormal proliferative disorder; cancer; skin tone;
KW  epidermal system; cosmetic improvement; skin tone; elasticity; zalphal;
KW  epitope.
XX
OS  Homo sapiens.
XX
PN  WO9929720-A2.
XX
PD  17-JUN-1999.
XX
PF  10-DEC-1998; 98WO-US26273.
XX
PR  10-DEC-1997; 97US-0987926.
XX
PA  (ZYMO ) ZYMOGENETICS INC.
XX
PI  Conklin DC, Lok S, Parish J;
XX
DR  WPI; 1999-385572/32.
XX
PT  Mammalian alpha helical protein-1, designated zalphal
XX
PS  Claim 10; Page 70; 73pp; English.
XX
CC  The present sequence is an epitope-bearing protein derived from
CC  human alpha helical protein-1, Zalphal. The protein comprises
CC  helices A, B, C and D of mature zalphal. It is used to raise specific
CC  antibodies which can be used for detection and purification of zalphal.
CC  The zalphal protein may be useful in the treatment of Fragile-X
CC  syndrome and abnormal proliferative disorders e.g. cancer. It can also
CC  be used for the growth, differentiation, maintenance and survival of
CC  connective tissues, particularly cardiovascular and epidermal systems
CC  and in imparting cosmetic improvements to normal connective tissues such
CC  as enhancement of skin tone and elasticity.
XX
SQ  Sequence 110 AA;

Query Match 7.0%; Score 149.5; DB 20; Length 110;
Best Local Similarity 27.3%; Pred. No. 0.00014;
Matches 36; Conservative 18; Mismatches 53; Indels 25; Gaps 1;
OY      103 NNNIOSOVOKITEENTLUREQVEPTPEDEDDIELRGAAAAAPPPPIEECPEDLPEKF 162
Db      1  nsqimeqlrllycerasillrvrp-----pscpvpifpet 35
OY      163 DGNNDMLAPFMAOCQIFEMKSTRDFSVDRVRCFVTSNMTGRARASAKLERSHYLMHN 222
Db      36  ngessrlpeflvtqasymlvnenfcdamkvaflislltgeeevvpylendspllgd 95

```

```

OY      223 YPAFMEMKHFV 234
Db      96  yrafldemkqcf 107

RESULT      9
AAV94674
ID  AAV94674 standard; Protein; 168 AA.
XX
AC  AAV94674;
XX
DT  01-DEC-2000 (first entry)
XX
DE  Human zsig83 mature protein sequence.
XX
KW  Alpha-helical protein; zsig83; cell growth; differentiation; cancer;
KW  Proliferation; chromosome 22q13.1-q13.2; cytostatic; vulneray;
KW  degenerative condition; metastasis; wound healing.
XX
OS  Homo sapiens.
XX
FH  Key
FH  Location/Qualifiers
FT  Domain
FT  /note= "SH3-binding domain"
FT  Region
FT  /label= Helix_A
FT  Region
FT  /note= "Hydrophillic region"
FT  Region
FT  /label= Helix_B
FT  Region
FT  /note= "Hydrophillic region"
FT  Region
FT  /label= Helix_D
FT  Region
FT  /note= "Hydrophillic region"
FT  Region
FT  /label= Helix_E
FT  Region
FT  /note= "Hydrophillic region"
XX
PN  WO200050594-A2.
XX
PD  31-AUG-2000.
XX
PF  25-FEB-2000; 2000MO-US04816.
XX
PR  26-FEB-1999; 99US-0259131.
XX
PA  (ZYMO ) ZYMOGENETICS INC.
XX
PI  Presnell SR;
XX
DR  WPI; 2000-572091/53.
DR  N-PDB; AAA28032.
XX
PT  Alpha-helical protein zsig83, its antibodies and the polynucleotide
PT  encoding the protein useful for treating disorders associated with
PT  abnormal cell growth e.g. cancer and agonists useful for treating
PT  wounds -
XX
PS  Claim 4; Page 75; 83pp; English.
XX
CC  This invention relates to a novel human alpha-helical protein designated
CC  zsig83. zsig83 plays a role in the process of cell growth,
CC  differentiation, or proliferation. The zsig83 gene is located on
CC  chromosome 22 at position 22q13.1-q13.2. Included in the invention are
CC  polynucleotide sequences encoding the zsig83 protein, expression vectors
CC  containing the zsig83 DNA sequence, a cultured cell containing the

```


XX 11-OCT-2001.
PD
XX 30-MAR-2001; 2001MO-US08631.
PF
XX 31-MAR-2000; 2000US-0540217.
PR
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PI
PI Drmanac RT, Liu C, Tang YT;
DR
DR WPI; 2001-639362/73.
DR
DR N-PSDB; AAF76392.
XX
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 20; SEQ ID NO 42564; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
XX Sequence 142 AA;
SQ
Query Match 6.5%; Score 139.5; DB 22; Length 142;
Best Local Similarity 31.4%; Pred. No. 0.0012;
Matches 36; Conservative 16; Mismatches 54; Indels 11; Gaps 3;
QY 129 EDEDDIELRCMAA-----AAPPEIEEC--PEDLPKFDGNDMLAFMAQCQIF 179
DB 20 eadsrdsdlrtvcagllqallalipratrwrnpfpctdgdtdrlpeflvtcsy 79
QY 180 MEKTRDPSPDVRVVCVTSMATGGAARMAKLEERSHYLMHNTPAFMEKRVF--EDP 237
DB 80 mivgenfssdalkvfltrltgpalqvwiplyikkesplndyrgflaemkrvfgedd 139
QY 238 Q 238
DB 140 e 140
RESULT 12
AAB60475
ID AAB60475 standard; Protein: 113 AA.
AC AAB60475;
XX
XX 24-APR-2001 (first entry)
DT
XX Human cell cycle and proliferation protein CCYPR-23, SEQ ID NO:23.

XX
XX Cell cycle and proliferation protein; CCYPR; human; agonist;
KM antagonist; gene therapy; detection; gene therapy;
KM transgenic animal disease model; immune disorder;
KM developmental disorder; cell signalling disorder;
KM cell proliferative disorder; cancer; tumour; anaemia; epilepsy;
KM arteriosclerosis; asthma; allergy; diabetes mellitus;
KM menstrual cycle disorder; bacterial infection.
XX
XX Homo sapiens.
OS
XX
XX WO200107471-A2.
PN
XX
XX 01-FEB-2001.
PD
XX
XX 21-JUL-2000; 2000MO-US19948.
PF
XX
XX 21-JUL-1999; 99US-0145075.
PR
XX 08-SEP-1999; 99US-0153129.
PR
XX 10-NOV-1999; 99US-0164647.
XX
XX (INCY-) INCYTE GENOMICS INC.
PA
XX
XX Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman O;
PI
PI Azimzal Y, Yang J, Lu DM, Baughn MR, Patterson C, Shah P;
XX
XX WPI; 2001-112727/12.
DR
DR N-PSDB; AAF59612.
XX
XX
XX Human cell cycle and proliferation proteins and polynucleotides are
PT used to treat, diagnose and prevent immune, developmental and cell
PT signalling disorders and cell proliferative disorders including cancer -
XX
XX Claim 1; Page 131; 205pp; English.
XX
XX Sequences AAB60453-AAB60506 represent 54 human cell cycle and
CC proliferation proteins (CCYPR), which are encoded by AAF59590-AAF59643.
CC CCYPR and agonists or CCYPR are used to treat diseases or conditions
CC associated with decreased expression of functional CCYPR, while CCYPR
CC antagonists are used to treat diseases or conditions associated with
CC overexpression of functional CCYPR. Monoclonal or polyclonal antibodies
CC to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or
CC radioimmunoassays to detect CCYPR. CCYPR itself may be used to detect
CC compounds e.g., antibodies, oligonucleotides and proteins (receptors)
CC that specifically bind to CCYPR, and in drug screening methods to
CC identify compounds that modulate the activity of CCYPR. CCYPR
CC nucleotides can be used to generate transgenic animal models of human
CC disease, and can be used in gene therapy in target cells with genetic
CC abnormalities with respect to the expression of CCYPR for the
CC treatment or prevention of a disorder associated with CCYPR.
CC Diseases which can be diagnosed, treated and prevented using CCYPR
CC proteins, nucleic acids, agonists or antagonists include immune,
CC developmental and cell signalling disorders, and cell proliferative
CC disorders including cancer. Specific examples of these disorders
CC include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,
CC diabetes mellitus, disorders of the menstrual cycle and infections
CC caused by bacteria.
XX
XX
XX Sequence 113 AA;
SQ
Query Match 6.5%; Score 139; DB 22; Length 113;
Best Local Similarity 35.0%; Pred. No. 0.0011;
Matches 28; Conservative 16; Mismatches 36; Indels 0; Gaps 0;
QY 155 PEDLPKFDGNDMLAFMAQCQIFMEKSTRDPSVDRVVCVTSMATGGAARMAKLE 214
DB 26 pipfepetfdgtdrlpeflvtcsymfvdcntfssdalkvfltrltgpalqvwiplyik 85
QY 215 RSHYLMHNTPAFMEKRVF 234
DB 86 kesplndyrgflaemkrvf 105

RESULT 13

AAVS9927
ID AAVS9927 standard; Protein; 144 AA.

AC AAVS9927;

DT 28-JAN-2000 (first entry)

DE Human myometrium tumour EST encoded protein 7.

KW Myometrium; tumour; human; expressed sequence tag; EST; uterine myoma;

KM treatment; carcinoma; cancer; gene therapy.

XX Homo sapiens.

XX DE19617947-A1.

XX 28-OCT-1999.

XX 17-APR-1998; 98DE-1017947.

XX 17-APR-1998; 98DE-1017947.

XX (META-) METAGEN GES GENOMFORSCHUNG MBH.

XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;

XX WPI; 1999-602380/52.

XX DR N-PSDB; AA241965.

XX New nucleic acid sequences expressed in uterine myoma, and derived

XX polypeptides, for treatment of uterine carcinoma and identification of

XX therapeutic agents

XX Claim 23; Page 68; 86pp; German.

CC This invention describes novel polypeptide sequences (I), fragments of
CC (I) fragments and their encoding nucleic acids (II) which are highly
CC expressed in human uterine myoma. (II) are used for recombinant
CC expression of (I) and to isolate complete genes. (I) are used to
CC identify agents suitable for treatment of uterine carcinoma, to directly
CC treat this form of cancer (including expression from gene therapy
CC vectors) and are used in a preparation for cancer treatment. (I) is also
CC used for the generation of specific antibodies. (II) are identified by
CC assembling ESTs (expressed sequence tags) from a particular tissue type
CC before comparison of expression patterns. This allows a significantly
CC longer fragment of the gene to be revealed and therefore reduces the
CC number of failures associated with the fact that ESTs from different
CC libraries may represent different parts of the same unknown gene.
CC disrupting the estimated frequency of occurrence in a particular tissue.
CC AAVS9921-Y99940 represent protein fragments encoded by the human
CC myometrium tumour cDNA library derived EST fragments represented in
CC AA241950-241980.

XX Sequence 144 AA;

SO

Query Match 6.4%; Score 137; DB 20; Length 144;
Best Local Similarity 29.4%; Pred. No. 0.002;
Matches 37; Conservative 19; Mismatches 58; Indels 12; Gaps 3;

QY 121 REVEPTPE-DEDDDIELKGA-----AAAAAPPPPIEEEC--PEDLPPEFGNPPM 168
DB 11 rrlisadphatqnsaaagymdgvrqlmkallagplrparrwrnpipifeftdgdtdr 70
QY 169 LAFPMACQIFEMKSTRDPSVDVRVRCFVTSMTGRRARASAKLDRSHYLMHNYPAFMW 228
DB 71 lpefiqtcsymfvdentfsndaklvtflitrltpalgwvipyikesplindygflia 130
QY 229 EMKRVF 234
DB 131 emkrvf 136

RESULT 14

AAW48895
ID AAW48895 standard; Protein; 1230 AA.

AC AAW48895;

DT 13-OCT-1998 (first entry)

DE Candida albicans CST20 protein.

KW CST20; protein kinase; Ste20p/p65PAK family; screening; virulence;

KM hyphal formation; pathogenic fungi; inhibitor; inflammation;

XX antimycotic.

XX Candida albicans.

XX W09818927-A1.

XX 07-MAY-1998.

XX 29-OCT-1997; 97WO-CA00809.

XX 30-OCT-1996; 96US-0029458.

XX (CANA) NAT RES COUNCIL CANADA.

XX Leberer E, Thomas DY;

XX WPI; 1998-272222/24.

XX DR N-PSDB; AAV32553.

XX In vitro screening test for agents that inhibit Candida genes

XX involved in virulence - and transition to hyphal form, potentially

XX useful as antimycotic agents

XX Disclosure; Fig 3; 79pp; English.

PS The sequence is that of the CST20 protein which can be used
XX in the development of an in vitro screening test for compounds
XX that inhibit biological activity of the protein and a system for
XX measuring its activity. The protein is involved in virulence and
XX hyphal formation. Inhibitors are potentially useful for rendering
XX pathogenic fungi (any species in which hyphal induction by kinase
XX occurs) avirulent and/or to treat inflammation.

SO Sequence 1230 AA;

Query Match 6.3%; Score 135.5; DB 19; Length 1230;
Best Local Similarity 22.5%; Pred. No. 0.043;
Matches 67; Conservative 54; Mismatches 114; Indels 63; Gaps 15;

QY 17 GGDPPGLHPRR-SEATAGRSPTPTVTLGDCPPPPPPNNNNNNNSKHTGKGA--- 72
DB 770 ggenalipqrlnefkahrappp--plappapvpappanllseqtselpqtrlapj 827
QY 73 -----CVPNTERRRDELSEELNLEKVKQSEENNNLOSQVOKLTENTTLRQVPR 126
DB 828 aladvteapnllylqgtkygeagqklrekkarelee-----lqrlreknecgnqget 880
QY 127 TPDEDDDIELKGA AAAAAPPPIEEECPEDDLPEKFDGND-MLAPPMAO----- 175
DB 881 ggnad-----taagsnlaiprvpvpnkp---psgs99grdakqaallqakkrekkkn 933
QY 176 COIFEMKST-----RDSVDVRVRCFVTSMTGRRARASAKLDRSHYLMHNYPAFMW 228
DB 934 lqtlaklktcpgdpneilydlvkl-----ggasgvtlanvdksniyai 982
QY 229 EMKRVFDPQRRVARRKIRLRQGM-GSVYIDYSAFOMIADLDWNEPALIDQYHEG 285
DB 983 kqmnlegqp-kkelinellvnmkslhpnlvntidsey-llkgdl-w-----vimeymeg 1033

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2002, 15:38:18 ; Search time 13.41 Seconds

(without alignments)
1157.833 Million cell updates/sec

Title: US-09-631-863a-2

Perfect score: 2139

Sequence: 1 MNMKRYLTKRRKRRSGRGOD.....DNCPAKAKSSPAGNSAPLV 401

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Database : SwissProt_40*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	159	7.4	146	LDOL_HUMAN	G95751 homo sapien
2	156	7.3	520	CET1_CANAL	O33803 candida alb
3	135	6.3	537	GAG_MLVVD	P11269 radiat ion m
4	131.5	6.1	538	GAG_MLVFE	P26806 friend muri
5	131.5	6.1	1230	STZD_CANAL	G92212 candida alb
6	130	6.1	537	GAG_MLVBA	P29167 murine leuk
7	129	6.0	872	S3B2_HUMAN	Q13433 homo sapien
8	122.5	5.7	536	GAG_MLVDE	P29168 murine leuk
9	122.5	5.7	538	GAG_MLVVD	P26805 friend muri
10	121.5	5.7	1585	P3K3_DICDI	P54675 dictyosteli
11	120.5	5.6	1157	Y182_HUMAN	Q14687 homo sapien
12	120.5	5.6	1816	AF6_HUMAN	P55196 homo sapien
13	120.5	5.6	1982	CHD12_DROME	O97159 drosophila
14	120	5.6	745	HG12_ARATH	P46607 arabidopsis
15	120	5.6	1905	TAGB_DICDI	P34683 dictyosteli
16	120	5.6	1937	MYH8_HUMAN	P13533 homo sapien
17	119.5	5.6	1939	MYH1_HUMAN	P12882 homo sapien
18	119	5.6	465	MYH6_RABIT	P04460 oryctolagus
19	119	5.6	505	WASL_HUMAN	O00401 homo sapien
20	118.5	5.5	428	INVO_CEBAL	P24709 cebus albif
21	118.5	5.5	2564	SPCO_HUMAN	G9H254 homo sapien
22	118	5.5	2175	HMCU_DROME	P10186 drosophila
23	117.5	5.5	501	WASL_RAT	O08816 rat mus musc
24	116.5	5.4	281	CEBE_RAT	P56261 rat mus musc
25	116.5	5.4	957	NEDA_MOUSE	P46933 mus musculi
26	116.5	5.4	2142	BAT2_HUMAN	P33336 homo sapien
27	116	5.4	537	GAG_MLVAV	O97592 canis fami
28	116	5.4	3680	DMD_CANFA	P27460 cas-br-e mu
29	115.5	5.4	536	GAG_MLVCB	P27460 cas-br-e mu
30	115	5.4	1227	B3A3_MOUSE	P36183 mus musculi
31	115	5.4	1332	SP77_YEAST	P15177 saccharomyc
32	115	5.4	1935	MYH7_HUMAN	P12883 homo sapien
33	115	5.4	1935	MYH7_PIG	P79293 sus scrofa

34	115	5.4	1939	MYH4_HUMAN	O9Y623 homo sapien
35	114.5	5.4	622	LAMO_DROME	P08928 drosophila
36	114.5	5.4	633	LA17_YEAST	Q12446 saccharomyc
37	114.5	5.4	1941	MYH2_HUMAN	G9UKX2 homo sapien
38	114	5.3	640	ELL2_HUMAN	O00472 homo sapien
39	113.5	5.3	281	CEBE_HUMAN	Q15744 homo sapien
40	113	5.3	1727	ALM1_SCHPO	G9UKX5 schistosach
41	113	5.3	1934	MYH7_MESAU	P13540 mesocricetu
42	112.5	5.3	2842	APC_RAT	P70478 rat mus musc
43	112	5.2	502	WASP_HUMAN	P42768 homo sapien
44	112	5.2	1223	YWRI_CAELN	Q10925 caenorhabdi
45	111.5	5.2	1232	B3A3_HUMAN	P48751 homo sapien

ALIGNMENTS

RESULT 1

ID LDOL_HUMAN STANDARD: PRT: 146 AA.

AC G95751;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE LDOL1 protein (leucine zipper protein down-regulated in cancer cells).

GN LDOL1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Fetal brain;

RA MEDLINE=9930357; PubMed=10403563;

RA Nagasaki K., Manabe T., Hanzawa H., Maass N., Tsukada T.,

RA Yamaguchi K.;

RT "Identification of a novel gene, LDOL1, down-regulated in cancer cell

RL lines ";

CC Cancer Lett. 140:227-234(1999).

CC -1- FUNCTION: MAY HAVE AN IMPORTANT ROLE IN THE DEVELOPMENT AND/OR

CC PROGRESSION OF SOME CANCERS.

CC -1- SUBCELLULAR LOCATION: Nuclear.

CC TISSUE SPECIFICITY: UNOUTGROSSLY EXPRESSED WITH HIGH LEVELS IN

CC BRAIN ANT THYROID AND LOW EXPRESSION IN PLACENTA, LIVER AND

CC LEUCOCYTES. EXPRESSED AS WELL IN SIX OF THE SEVEN HUMAN BREAST

CC CANCER CELL LINES EXAMINED.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

CC EMBL; AB019527; BAA34364.1; -

DR EMBL; AB019527; BAA34364.1; -

KW Nuclear protein.

FT DOMAIN 132 143 ASP/GLU-RICH (HIGHLY ACIDIC).

SO SEQUENCE 146 AA; 16968 MW; 3D02813B2DE52DBE CRC64;

Query Match 7.4%; Score 159; DB 1; Length 146;

Best local similarity 27.4%; Pred. No. 0.0007;

Matches 45; Conservative 21; Mismatches 68; Indels 30; Gaps 2;

QY 83 DELSEINNLREKVKQSENNNNLOSOVKLTREQVEPTPEDEDDIELRGAA 142

DB 3 DELVLLALLALMRKALSIENSQMLEQRLVCEASLLRQYRP----- 46

QY 143 AAAPPPLEECPEDDLPKFGDNPMILAPFAAQOIFEMKSTRFSDVRVCFVTSMT 202

DB 47 -----PSCPVPFPPTFNGESSRLPEFLVOTFASVYLVNENRCNDAMKVAFLISLLT 97

FT LIPID 2 2 MYRISTATE.
 FT CONFLICT 479 479 T -> S (IN REF. 2).
 FT CONFLICT 482 482 T -> S (IN REF. 2).
 SQ SEQUENCE 537 AA; 60784 MW; 312AF7B2BB4B7FB CRC64;

Query Match 6.3%; Score 135; DB 1; Length 537;
 Best Local Similarity 19.1%; Pred. No. 0.11; Indels 180; Gaps 20;

Matches 97; Conservative 58; Mismatches 173; Indels 180; Gaps 20;

QY 21 PGLHPRS-----EATAGRSPP-----TPVTYLGPDGP-----PPPPPPPPNNNN 61
 DB 71 PGPCHHPDQVPYIVTWEAIAVEPPSWVKPFSKLSLSTPAIILSGSTQPP----- 124
 QY 62 NNSKTKHKSACVPMTERRDELSEIINLREKYMKOSEENNUNLOSOYQKLTENTTLR 121
 DB 125 -----RSALYPALTP-----SIKPRSKQVILSDNGCPRLDLTETDPPPYG 165
 QY 122 EQVEPTPEDEDDIELRGAAAAAAP-----PPPIEECPEDLPKFDGNPD- 168
 DB 166 EQGPSSPGDGDREATYTSSEIPAPSPVSRIGRKRPADSTSRAPFLRLGNGQLQ 225
 QY 169 LAPFAQCOIFMEKSTDFSDVDRVY-CFVYSMTGGAARWASAKLERSHYLMHYPAFM 227
 DB 226 YWPFSSDLYNWKNNPFSFEDPGKLTALIESVLTTHQPTWDDCOQ-----L 272
 QY 228 MEMKIVFEDPQREYAKKIRRLRQMGSVI--DYSNAFQMTAODLDNMP----- 276
 DB 273 LGTLLTGEEKORVILEARKAVANGDGRPTQLPNEVNSAFPLERPMDYTTTPEGRNHLVLY 332
 QY 277 -----ALIDYHEGL-----SDHIQEE- 292
 DB 333 ROLLLAGIOMNCRPTNLAKVGTGPNESPFAFLERLKEAYRRYTPDPEDHGOETSV 392
 QY 293 -----ELSHLE--VAKSLSALIGQCIHT-----E 314
 DB 393 SMSFTIQSAPDIGRKLERLEDLSKTLIDLVEAREKIFMKRRETPREBERFRRETEEME 452
 QY 315 RRLAATAAARRPRPPR-----ALVLPHTASHHQVPTPEVGARMLTQEEKERRKRL 370
 DB 453 RRRADEQREKERDRRRREMSKLLATVYTGORODRO--GGERRR-PQLDKDQ----- 502
 QY 371 CLYCGTGGHYADNCPAKASKSSPAGNSP 398
 DB 503 CAYCKEKHMAKDCPKK--PGCPRGPRP 528

RESULT 4
 GAG_MLVFF
 ID GAG_MLVFF STANDARD: PRT: 538 AA.
 AC P26806;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE GAG polypeptide (Core polypeptide) [Contains: Matrix protein P15; RNA binding phosphoprotein P12; Capsid protein P30; Nucleocapsid protein P10].
 DB P10.
 GN GAG.
 OS Friend murine leukemia virus (isolate FB29) (F-MuLV).
 OC Viruses; Retroviral viruses; Retroviridae; Mammalian type C retroviruses.
 OX NCBI_TaxID=11797;
 RN 11
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92107687; PubMed=1762923;
 RA Perryman S., Nishio J., Chesebro B.;
 RT "Complete nucleotide sequence of Friend murine leukemia virus, strain FB29".
 RL Nucleic Acids Res. 19:6950-6950(1991).
 RN 12
 RP SEQUENCE OF 479-512.
 RX MEDLINE=81264245; PubMed=6267042;
 RA Henderson L.E., Copeland T.D., Sowder R.C., Smythers G.W.,
 ROroszian S.;

RT "Primary structure of the low molecular weight nucleic acid-binding proteins of murine leukemia viruses.";
 RL J. Biol. Chem. 256:8400-8406(1981).
 CC -I- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC -I- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-POLYPROTEIN.
 CC -I- SIMILARITY: CONTAINS 1 CCHC-TYPE ZINC FINGER.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sdb.ch/announce/> or send an email to license@isb-sdb.ch).
 CC EMBL; 211128; CAA7478.1; -
 DR InterPro: IPR000840; Gag_MA.
 DR InterPro: IPR002079; Gag_P12.
 DR InterPro: IPR003036; Gag_P10.
 DR InterPro: IPR001878; Znf_CCHC.
 DR Pfam: PF01140; gag_MA; 1.
 DR Pfam: PF01141; gag_P12; 1.
 DR Pfam: PF02093; Gag_P30; 1.
 DR Pfam: PF00098; ZF-CCHC; 1.
 DR SMART: SM00343; ZNF_C2HC; 1.
 DR PROSITE: PS0158; ZF_CCHC; 1.
 KW Coat protein; Core protein; Polypeptide; Nucleoprotein; Myristate; Phosphorylation; Zinc-finger.
 FT CHAIN 2 131
 FT CHAIN 132 215
 FT CHAIN 216 478
 FT CHAIN 479 538
 FT ZN_FING 502 519
 FT LIPID 2 2
 FT CONFLICT 501 501
 SQ SEQUENCE 538 AA; 60929 MW; 2E652DD9B9E0DA3C CRC64;

Query Match 6.1%; Score 131.5; DB 1; Length 538;
 Best Local Similarity 18.8%; Pred. No. 0.18;
 Matches 95; Conservative 53; Mismatches 185; Indels 171; Gaps 17;

QY 21 PGLHPRS-----EATAGRSPP-----TPVTYLGPDGP-----PPPPPPPPNNNN 64
 DB 71 PGPCHHPDQVPYIVTWEAIAADPPWVPRFVHKKPPLSLPSPAPSLPPPPPLS----- 123
 QY 65 KHTGHSACVPMTERRDELSEIINLREKYMKOSEENNUNLOSOYQKLTENTTLR 124
 DB 124 -TPQSSLYPALT-----SPLWTKRRPVYLPDSG-----GPLDLTETDPPPYRDPG 169
 QY 125 EPTPEDEDDIELRGAAAAAAP-----PPPIEECPEDLPKFDGNPD-MUAP 171
 DB 170 PPSPGNGSGEVALTEGAPDPSPVSRIGRKRPADSTSQAFPLRLGNGQYQWYP 229
 QY 172 FMAQCOIFMEKSTDFSDVDRVY-CFVYSMTGGAARWASAKLERSHYLMHYPAFM 230
 DB 230 FSSSDLYNWKNNPFSFEDPAKLTALIESVLTTHQPTWDDCOQ-----LLGT 276
 QY 231 KHVFDQREYAKKIRRLRQMGSVI--DYSNAFQMTAODLDNMP----- 274
 DB 277 LITGEEKORVILEARKAVANGDGRPTQLPNDINDAFLEPRPMDYNTQGRNHLVHYRQL 336
 QY 275 -----EPALIDYHEGL----- 286
 DB 337 LLAGIOMNCRPTNLAKVGTGPNESPFAFLERLKEAYRRYTPDPEDPGQETVAMS 396
 QY 287 -----SDHIQEEELSHLE--VAKSLSALIGQCIHT-----ERRL 317
 DB 397 FIMQSPADIGRKLERLEDLSKTLIDLVEAREKIFMKRRETPREBERIRRETEKEERRR 456
 QY 318 ARAAARPRPRPALVLPHTASHHQVPTPEVGARMLTQEEKERRKRL--NLCLYC 374

Db 457 AEDVQREKERDRRR-----HREMSKLATVVSQRODROGERRRRPOLDHDOCAVC 507
 QY 375 GTGGHVADNCPAKASKSPAGNSP 398
 ||:|||||
 Db 508 KEGHWARDCPK--PRGPRPP 529

RESULT 5
 ST20_CANAL STANDARD; PRT; 1230 AA.
 AC 092212;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Serine/threonine-protein kinase STE20 homolog (EC 2.7.1.-).
 GN HST20 OR CSTR20.
 OS Candida albicans (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
 OX NCBI_TaxID=5476;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1066; PubMed=8917572;
 RA Medline=97075146; Fink G.R.;
 RT "Candida albicans strains heterozygous and homozygous for mutations
 in mitogen-activated protein kinase signaling components have defects
 in hyphal development.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:13223-13228(1996).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC STE20 SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS 1 GBD DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC
 DR EMBL: U73457; AAB38875.1; -.
 DR HSSP: P00518; 1PHK.
 DR Interpro: IPR000719; Euk_pkinase.
 DR Interpro: IPR000095; PAK_box_P21_Rho_binding.
 DR Interpro: IPR002290; Ser_thr_kinase.
 DR Pfam: PF00786; PBD; 1.
 DR Pfam: PF00069; pkinase; 1.
 DR SMART: SM00285; PBD; 1.
 DR SMART: SM00220; S_TKC; 1.
 DR PROSITE: PS50108; GBD; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 KW transferase; Serine/threonine-protein kinase; ATP-binding.
 FT DOMAIN 109 119
 FT PIR: A40416; FOMVMB.
 FT Interpro: IPR000840; Gag_MA.
 FT Interpro: IPR002079; Gag_P12.
 FT Interpro: IPR003036; Gag_P30.
 FT Interpro: IPR001878; Znf_CCHC.
 FT Pfam: PF01140; gag_MA; 1.
 FT Pfam: PF01141; gag_P12; 1.
 FT Pfam: PF02093; gag_P30; 1.
 FT Pfam: PF00098; zfc_CCHC; 1.
 FT SMART: SM00343; Znf_C2HC; 1.
 DR PROSITE: PS50158; Zf_CCHC; 1.
 KW Coat protein; Core protein; Polyprotein; Nucleoprotein; Myristate;
 ZINC-finger.
 FT CHAIN 2 129 CORE PROTEIN P15.
 FT CHAIN 130 214 INNER COAT PROTEIN P12.
 FT CHAIN 215 477 CORE SHELL PROTEIN P30.

Query Match 6.1%; Score 131.5; DB 1; Length 1230;
 Best Local Similarity 21.5%; Pred. No. 0.46;

Matches 64; Conservative 51; Mismatches 121; Indels 61; Gaps 13;
 QY 17 GGDPGLPHPR-SEATAGSPPTVTVLGDPCEPPPPPPNNNNNNSKHTGKSA--- 72
 ||:|||||
 Db 770 GGNNALPKRINFENKRRAPPP--PSAPRAPPVPAPANLLSEQTSILPOQRTAPSQ 827
 QY 73 -----CVPNTERRDELSEINILREKVMKSENNNISOYOKITEENTILREOYEP 126
 ||:|||||
 Db 828 ALADVTAPTNIVEIQOTKYEOAOOKLEKKARELEE-----IORLEKNERONROQET 880
 QY 127 TPEDEDDIELRGAAAAPPPPEECPEDELPEKFGNDP-MLAPMAO----- 175
 ||:|||||
 Db 881 GQNNAD-----TASGSGINAPVPVPNKKP--PSGSGGDAQALIAQKREKKRN 933
 QY 176 QQIMEKST-----RDFSVDRYRCFVTSMTGARAARASAKLESHYLMHYPAFM 228
 ||:|||||
 Db 934 LQITAKLKTICNPDPRELVDLKI-----GQASGGVFLAHVDRKSNVAL 982
 QY 229 EMKHVEFDPQREYAKKIRLRQMGSVIYDSNAFQIMQDLDMNEPALIDYHEG 285
 ||:|||||
 Db 983 KQNLLEQOPKKEILLINELVMKSSHPNIVNFTDSY-LLKGDL-W---VIMEYMEG 1033

RESULT 6
 GAG_MLYBM STANDARD; PRT; 537 AA.
 ID GAG_MLYBM
 AC P29167;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Gag polyprotein [contains: Core protein p15; Inner coat protein p12;
 DE Core shell protein p30; Nucleoprotein p10].
 GN Gag.
 OS Murine leukemia virus (strain BM5 eco).
 OS Viruses; Retroid viruses; Retroviridae; Gammaretrovirus.
 OX NCBI_TaxID=31687;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91303677; PubMed=1649328;
 RA Chattopadhyay S.K., Sengupta D.N., Fredrickson T.N., Morse H.C. III,
 RA Hartley J.W.;
 RT "Characteristics and contributions of defective, ecotropic, and mink
 cell focus-inducing viruses involved in a retrovirus-induced
 RT Immunodeficiency syndrome of mice.";
 RL J. Virol. 65:4232-4241(1991).
 CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC -1- SIMILARITY: CONTAINS 1 CCHC-TYPE ZINC FINGER.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC
 DR EMBL: M64095; AAA46510.1; -.
 DR PIR: A40416; FOMVMB.
 DR Interpro: IPR000840; Gag_MA.
 DR Interpro: IPR002079; Gag_P12.
 DR Interpro: IPR003036; Gag_P30.
 DR Interpro: IPR001878; Znf_CCHC.
 DR Pfam: PF01140; gag_MA; 1.
 DR Pfam: PF01141; gag_P12; 1.
 DR Pfam: PF02093; gag_P30; 1.
 DR Pfam: PF00098; zfc_CCHC; 1.
 DR SMART: SM00343; Znf_C2HC; 1.
 DR PROSITE: PS50158; Zf_CCHC; 1.
 KW Coat protein; Core protein; Polyprotein; Nucleoprotein; Myristate;
 ZINC-finger.
 FT CHAIN 2 129 CORE PROTEIN P15.
 FT CHAIN 130 214 INNER COAT PROTEIN P12.
 FT CHAIN 215 477 CORE SHELL PROTEIN P30.

FT CHAIN 478 537 NUCLEOPROTEIN P10.
 FT 2N.FING 501 518 CCHC-TYPE.
 FT LIPID 2 MYRISTATE (BY SIMILARITY).
 SO SEQUENCE 537 AA: 60422 MW: AAD2E70299BFFD64 CRC64.

Query Match 6.1%; Score 130; DB 1; Length 537;
 Best Local Similarity 18.5%; Pred. No. 0.22;
 Matches 93; Conservative 53; Mismatches 186; Indels 172; Gaps 17;

QY 21 POLAPHR-----EATAGRSP-----TPVTTLGPDGP-----PPPPPPNNNNN 61
 DB 71 PEPHGHQPOVPIYVWEIAYEPPEVVKPFVSPKLSLSTAPLIPSGSTQPP----- 124
 QY 62 NNSKRTGKSAQVPMTERRDELSEELNNLRKWKAKOSENNNNQSOYOKITEENTILR 121
 DB 125 -----RSALYPAFT-----PSIKRPSKPOVLSDDGPI-----IDLTEPPPPYG 165
 QY 122 EOVEPTPEDEDDILRGAAAAAP-----PPPIEECPEDPEKFGDGNPM- 168
 DB 166 EOGSPSPGDDGREATSTSEIPAPSPVSLRGKRDPADSTTSKAPPLRLGNGQLQ 225
 QY 169 LAFMAQOQIFMEKSTRDPSVDVAVR-CFVTSMTGRARAASAKLERSHYLMHNPAPFM 227
 DB 226 YWPFSSDLYMKNKNNPSFSDPGKLTALIESVLTTHQPTWDCQ-----L 272
 QY 228 MEMKIVFDPQREYAKRKIRRLROGMSVT--DYSNAPQMTAQLDWNENP----- 276
 DB 273 LGTLTGEKKQVLLERAKAVANGNGRPTQLPNEVNSAFLEPRPDMDYTTPEGRNHLVLY 332
 QY 277 -----ALIDQHELSH----- 289
 DB 333 ROLLLAGIOMNGRSPTNLAKYKGTGQNPESAFLEKLEKAYRRYPTYPDPDGOETNV 392
 QY 290 -----IOEELSHLE--VAKSLALIGOCIH-----E 314
 DB 393 SMSFTWQAPALGRLELEDLKSXTLGDIVREAKIFNKRTPERERIRIRETEKEE 452
 QY 315 RRLAARAAAKRRSPRALVLEPHIASHHQVDTEPVGAGRMRLTOEKEKRRKMLCLYC 374
 DB 453 RRRADDEOREKERDRRRRREMSKLLATVVTGQROROGGERRRRPLDQK-----CAVC 506
 QY 375 GTGHHYADNCPAKAKSSSPAGNSP 398
 DB 507 KEKGHWAKDCPKK--PRGPRGPRP 528

RESULT 7
 S3B2_HUMAN STANDARD: PRT: 872 AA.
 ID S3B2_HUMAN
 AC 013435;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Splicing factor 3B subunit 2 (Spliceosome associated protein 145) (SAP 145) (SF3B150) (Pre-mRNA splicing factor SF3B 145 Kda subunit).
 GN SF3B2 OR SAP145.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homalidae; Homo.
 NC NCBL_TaxID:9606;
 RX MEDLINE=96154048; PubMed=8566756;
 RA "Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A.";
 RL Genes Dev. 10:233-243(1996).
 RN [2]
 RP CHARACTERIZATION OF THE SPLICEOSOME.
 RP MEDLINE=20337962; PubMed=10882114;
 RA Das R., Zhou Z., Reed R.;

RF "Functional association of U2 snRNP with the ATP-independent
 RF spliceosomal complex E.";
 RL Mol. Cell 5:779-787(2000).
 CC -I- FUNCTION: SUBUNIT OF THE SPLICING FACTOR SF3B REQUIRED FOR 'A' COMPLEX ASSEMBLY FORMED BY THE STABLE BINDING OF U2 SNRNP TO THE BRANCHPOINT SEQUENCE (BPS) IN PRE-MRNA. SEQUENCE INDEPENDENT BINDING OF SF3A/SF3B COMPLEX UPSTREAM OF THE BRANCH SITE IS ESSENTIAL. IT MAY ANCHOR U2 SNRNP TO THE PRE-MRNA. MAY ALSO BE INVOLVED IN THE ASSEMBLY OF THE 'E' COMPLEX. BELONGS ALSO TO THE MINOR U12-DEPENDENT SPLICOSOME, WHICH IS INVOLVED IN THE SPLICING OF RARE CLASS OF NUCLEAR PRE-MRNA INTRON.
 CC -I- SUBUNIT: COMPONENT OF SPLICING FACTOR SF3B WHICH IS COMPOSED OF FOUR SUBUNITS: SF3B4/SAP49, SF3B3/SAP130, SF3B2/SAP145, SF3B1/SAP155. SF3B ASSOCIATES WITH THE SPLICING FACTOR SF3A AND A 12S RNA UNIT TO FORM THE U2 SMALL NUCLEAR RIBONUCLEOPROTEINS COMPLEX (U2 SNRNP). SF3B2 INTERACTS DIRECTLY WITH SF3B4.
 CC -I- SIMILARITY: TO YEAST CUS1. SOME, TO C.ELEGANS ZK632.11.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: U41371; AAA97461.1; .
 DR MIM: 605591; .
 DR InterPro: IPR003034; SAP.
 DR Pfam: PF02037; SAP; 1.
 DR SMART: SM00513; SAP; 1.
 DR Spliceosome; mRNA processing; mRNA splicing; Nuclear protein.
 KW DOMAIN 68 73
 FT DOMAIN 68 73
 FT DOMAIN 81 90
 FT DOMAIN 106 109
 FT DOMAIN 226 230
 FT DOMAIN 269 274
 FT DOMAIN 308 312
 FT DOMAIN 676 679
 FT DOMAIN 697 703
 FT DOMAIN POLY-GLU.
 SO SEQUENCE 872 AA: 97656 MW: AED669FDDDA5DE31 CRC64;

Query Match 6.0%; Score 129; DB 1; Length 872;
 Best Local Similarity 20.7%; Pred. No. 0.45;
 Matches 83; Conservative 64; Mismatches 150; Indels 104; Gaps 19;

QY 16 RGGDPC-----LHPRSEATAGRSPTPTV-TLGDPCPP--PPPPPPNNNNNSK 65
 DB 182 RPPQDMGQIGVETPIGP-RVAAPVGPVGPTPTVLWGAAPVPPRPPPPGDNENEMDP 240
 QY 66 HGHGSACVPMNTER---RDELSEELNNLRKWKAKOSENNNNQSOYOKITEENTILRE 122
 DB 241 SVGPR--IPQALEKIILQKESQDEEMNSQDEEMETDARSLSQASFTEDIVYSVK 297
 QY 123 Q-----VEPTPEDEDDILRGAAAAAPPPPIEECPEDLPE 160
 DB 298 KEKNKRRNRKRRKKRQVARGVSSGDRKSDRSRSKSDSPA---DVEIEVYTEPE 354
 QY 161 KFDGN-----PKLAFPMACQIFMEKSTRDPSVDVAVRCFVTSMTGGAARASAKLERS 216
 DB 355 IYEPNFIFFKRIEFAFKLTDVKKEREKPEKLD-----KLENS 393
 QY 217 HYLMIYPAFMEKHVEP-----PPREAKRKIRRLROGMSVIDYSNAPQMI 267
 DB 394 --AAKKKGFEENKDDSDSDQDEKKRPAKSKKKLRMMN-----FTVALKQLV 446
 QY 268 AODLDWNEPALIDYHESGISHIOEELSHLEVAKSLALIGOCIHIERLARAARPR 327
 DB 447 AR-----PDVVEHMDVTAOD--PKLVHLKATRN-SYVPRHMGCFKKRYLD--GKRGIE 495
 QY 328 SPRLALVLEPHIASHHQVDTEPVGAGRMRLTOEKEKRRKRL 368

```

Db 496 KPP--FELP-----DFIKRTGIOEMREALOEKEOKTM 526

RESULT 8
GAG_MLVE STANDARD; PRT; 536 AA.
ID GAG_MLVE
AC P29168;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE GAG polypeptide [contains: Core protein p15; Inner coat protein p12;
DE Core shell protein p30; Nucleoprotein p10].
GN GAG.
OS Murine leukemia virus (strain DEF27).
OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus.
NC NCBI_TaxID=31688;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91303677; PubMed=1649328;
RA Chattopadhyay S.K., Sengupta D.N., Fredrickson T.N., Morse H.C. III,
RA Hartley J.W.;
RT "characteristics and contributions of defective, ecotropic, and mink
RT cell focus-inducing viruses involved in a retrovirus-induced
RT immunodeficiency syndrome of mice.";
RT J. Virol. 65:4232-4241(1991).
CC -1- P1M: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 CCHC-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M64096; AAA46509.1; ALT_SEQ.
DR PIR: B40416; FOMVME.
DR InterPro: IPR000840; Gag_MA.
DR InterPro: IPR002079; Gag_P12.
DR InterPro: IPR003036; Gag_P30.
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam: PF01140; gag_MA; 1.
DR Pfam: PF01141; gag_P12; 1.
DR Pfam: PF02093; gag_P30; 1.
DR Pfam: PF00098; zf_CCHC; 1.
DR SMART: SM00343; Znf_C2HC; 1.
DR PROSITE: PS50158; Zf_CCHC; 1.
KW Coat protein; Core protein; Polypeptide; Nucleoprotein; Myristate;
KW Zinc-finger.
FT CHAIN 2 129 CORE PROTEIN P15.
FT CHAIN 130 213 INNER COAT PROTEIN P12.
FT CHAIN 214 476 CORE SHELL PROTEIN P30.
FT CHAIN 477 536 NUCLEOPROTEIN P10.
FT ZN_FING 500 517 CCHC-TYPE.
FT LIPID 2 2 MYRISTATE (BY SIMILARITY).
SQ SEQUENCE 536 AA; 60890 MW; 4D990FDC5AEF39B CRC64;

Query Match 5.78; Score 122.5; DB 1; Length 536;
Best local Similarity 18.84; Pred. No. 0.65;
Matches 97; Conservative 45; Mismatches 178; Indels 195; Gaps 17;

OY 21 PGLPHRS-----DATGRSP-----PTPTVTLGPPCCPPPPPPPPPPNNNNN 62
Db 71 PGPBGHDQVYIVTWALYHPPWKKPFVSPKPFPLSTL-PPSPGSAHPPS----- 124
OY 63 NSKHTGSKACVPMTERRDELSEETINLREKVKQSENNNLQSOVKLTETNTLRE 122
Db 125 -----RSDLYTALLPSITKPKPSRVLPFTNGGLIDLITENPNLIGE 166
OY 123 QVEPTPEDEDDIELRGAAAAAAPPIIEECPEDLPEKFDGNDMLA----- 170

```

```

Db 167 QGPPLP-----KSPVKKRRPPPP-RYSPPEPVSRLLGNDDPPAADSTTSRAEPLR 216
OY 171 -----PEMAQCCIFMEKSTRDFSDVRVY-CVTSMMTGRAARMAKLESHYL 219
Db 217 LGNGQLQIYMPFSSSDLYNKKNNPFSFEDPGKLTALIESVLTTHQFTWDDCO----- 270
OY 220 MNPAPFMEKMHVEDDPRREKVRIRRLRGQMSVI--DYSNAPQMIADIDMNEP- 276
Db 271 -----LLGTLTGEKQRYLLEARKAVRGNDRPQLPNEVNSAPPLERPDMDYTPPE 323
OY 277 -----ALDQYHEGL----- 286
Db 324 GRNHLVLYROLLAGLONAGRSPTNLAKYGTGGPSPSPAFLERLKEAYRRYTPDPE 383
OY 287 -----SDHIOEELSHLE--VAKSLALIGCINI----- 313
Db 384 DPGQETVNSMSEIWSAPDGRKLERLEDLSKTLGDIYREAEKIFNKRTEPERERIR 443
OY 314 -----ERLARAARAKRPPRPALVLPPIASHQVDPTEPYGARMRLTOEKEERR 366
Db 444 RETEKEERRRRADEQREKERDRR-----HREMSKFLATVTVGRODROGGERRR 494
OY 367 ---KLNLCITCGTGCHYADNCPAAKSSPAGNSP 398
Db 495 POLDKDCAYCEKRGHWAKDCPKR--PRGPRGPR 527

RESULT 9
GAG_MLVEP STANDARD; PRT; 538 AA.
ID GAG_MLVEP
AC P26805;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE GAG polypeptide (Core polypeptide) [contains: Matrix protein p15; RNA
DE binding phosphoprotein p12; Capsid protein p30; Nucleocapsid protein
DE p10].
GN GAG.
OS Friend murine leukemia virus (isolate PVC-211) (F-MuLV).
OC Viruses; Retroviral viruses; Retroviridae; Mammalian type C retroviruses.
NC NCBI_TaxID=11798;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92319660; PubMed=1620621;
RA Remington M.P., Hoffman P.M., Ruscetti S.K., Masuda M.;
RT "Complete nucleotide sequence of a neurotropic variant of Friend
RT murine leukemia virus PVC-211."
RT Nucleic Acids Res. 20:3249-3249(1992).
CC -1- P1M: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-POL
CC POLYPEPTIDE.
CC -----
CC -1- SIMILARITY: CONTAINS 1 CCHC-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M93134; AAA46476.1; -.
DR PIR: S35474; S35474.
DR InterPro: IPR000840; Gag_MA.
DR InterPro: IPR002079; Gag_P12.
DR InterPro: IPR003036; Gag_P30.
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam: PF01140; gag_MA; 1.
DR Pfam: PF01141; gag_P12; 1.
DR Pfam: PF02093; gag_P30; 1.
DR Pfam: PF00098; zf_CCHC; 1.
DR SMART: SM00343; Znf_C2HC; 1.

```

DR PROSITE: PS50158; 2F CCHC; 1.
 KW Coat protein; Core protein; Polypeptide; Nucleoprotein; Myristate;
 KM Phosphorylation; Zinc-finger.
 FT CHAIN 2 131 MATRIX PROTEIN P15.
 FT CHAIN 132 215 RNA BINDING PHOSPHOPROTEIN P12.
 FT CHAIN 216 478 CAPSID PROTEIN P30.
 FT CHAIN 479 538 NUCLEOCAPSID PROTEIN P10.
 FT 2N_FING 502 519 CCHC-TYPE.
 FT LIPID 2 MYRISTATE (BY SIMILARITY).
 SQ SEQUENCE 538 AA; 61033 MM; 2FF9F97D2C79DEBE CRC64;

Query Match 5.7%; Score 122.5; DB 1; Length 538;
 Best Local Similarity 18.6%; Pred. No. 0.65;
 Matches 98; Conservative 49; Mismatches 164; Indels 215; Gaps 20;

QY 21 POLHPRS-----EATGRSP-----PTPTVLGPPCPPPPPPNNNNNNNS 64
 DB 71 PGPCHHPQVPIYVWELAVDPFVHKRPDLPLPAPSLPPEPLIS----- 123
 QY 65 KHTGKSAQVPMTERRDELSSEINLRKYMKOSENNNLOSQYKITEENTTLREGV 124
 DB 124 -TPQSSLYPALT-----SPLNTKRPQVLPDSG-----GPLIDLTFEDPPPYRDPG 169
 QY 125 EPTPEDEDDIELRGAAAAAP-----PPPIEECPEDLPKFDGND-ML 169
 DB 170 PPSPGKCDGSEV--APTEGAPDSSPMVSRKGRREPPVADSTTSQAPFLRLGNGQFOY 227
 QY 170 APFMAOQIFMEKSTRDSVDVRY--CEVTSMGTGRAARMAKLESHYLMHNPAPFM 228
 DB 228 WPFSSSDLYNMKNNNPSSSEDPGKTLALIESVLTLHPTWDDCQ-----LL 274
 QY 229 EKHAFEPQREVAKKIRLRQMGSVY--DYSNAFQMLADIDMNEPALIDYHEGL 286
 DB 275 GTLLGTEKQRYLLEARKAVRGEDEGRPTOLPINDAPLEPDMDYN----- 322
 QY 287 SDHIOEESHLEVAKSLALIGQCIHERRLRAAARSPSPALVPHIAS----- 340
 DB 323 ---TORGNHL-----VHY-RQLLAGLQNNAGRSPTMLAKYKGTGQPNESP 365
 QY 341 -----HHQVDPTEP-----VGGRNRL----- 357
 DB 366 SAFELRLKEAVRYTPYDEPQGETNVSMSPFQWQAPDGLKRLLEDLAKKTLGDLVR 425
 QY 358 -----TOSEKERR-----KL----- 368
 DB 426 EAEKLFNKRREPEREREVRRETEKEEREREDEREKRRRRREMSKLLATVVSQ 485
 QY 369 -----NLCYCGTGHYADNCPAKASKSPAGNSP 398
 DB 486 RODRGGERRRPOLDHQCAVCKEKGHWARDCPKR--PRGPRGPRP 529

RESULT 10
 P3K3_DICDI STANDARD; PRT; 1585 AA.
 AC P54675;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Phosphatidylinositol 3-kinase 3 (EC 2.7.1.137) (PI3-Kinase)
 DE (Ptdins-3-kinase) (PI3K) (Fragment).
 GN PI3K OR PI3K.
 OS Dictyostelium discoideum (Slime mold).
 CC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
 OX NCBI_Taxid:44689;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-AX3;
 RX MEDLINE-96009592; PubMed-7565716;
 RA Zhou K., Takegawa K., Emr S.D., Firtel R.A.;
 RT "A phosphatidylinositol (PI) kinase gene family in Dictyostelium
 discoideum: biological roles of putative mammalian p110 and yeast

RT Vps34p PI 3-kinase homologs during growth and development."
 RL Mol. Cell. Biol. 15:5645-5656(1995).
 CC -I- CATALYTIC ACTIVITY: ATP + 1-phosphatidyl-ID-myo-inositol = ADP +
 CC 1-phosphatidyl-ID-myo-inositol 3-phosphate.
 CC -I- SIMILARITY: BELONGS TO THE PI3/PI4-KINASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-slb.ch/announce/>
 CC or send an email to license@isb-slb.ch).
 CC -----
 CC EMBL: U23478; AA85723.1; -.
 DR Dictydb; DD01101; PI3K.
 DR InterPro; IPR002420; PI3K_C2.
 DR InterPro; IPR000341; PI3K_ras_bind.
 DR InterPro; IPR001263; PI3K.
 DR InterPro; IPR00403; PI3_P14_kinase.
 DR Pfam; PF00613; PI3K; 1.
 DR Pfam; PF00792; PI3K_C2; 1.
 DR Pfam; PF00794; PI3K_rbd; 1.
 DR Pfam; PF00454; PI3_P14_kinase; 1.
 DR SMART; SM00142; PI3K_C2; 1.
 DR SMART; SM00144; PI3K_rbd; 1.
 DR SMART; SM00145; PI3K; 1.
 DR SMART; SM00146; PI3K; 1.
 DR PROSITE; PS00915; PI3_4_KINASE_1; 1.
 DR PROSITE; PS00916; PI3_4_KINASE_2; 1.
 DR PROSITE; PS00915; PI3_4_KINASE_3; 1.
 DR TRANSFERASE; Kinase; Multigene family; Repeat.
 KM NON_TER 1
 FT DOMAIN 1 84 POLY-ASN.
 FT DOMAIN 98 110 POLY-ASN.
 FT DOMAIN 200 226 POLY-ASN.
 FT DOMAIN 239 254 POLY-ASN.
 FT DOMAIN 345 378 POLY-ASN.
 FT DOMAIN 383 390 POLY-ASN.
 FT DOMAIN 484 488 POLY-SER.
 FT DOMAIN 720 737 POLY-GLN.
 FT DOMAIN 1569 1575 POLY-ASN.
 FT DOMAIN 1221 1484 POLY-GLN.
 FT DOMAIN 1510 1539 5 X 5 AA APPROXIMATE REPEATS.
 FT REPEAT 1510 1514 1.
 FT REPEAT 1515 1519 2.
 FT REPEAT 1520 1524 3.
 FT REPEAT 1530 1534 4.
 FT REPEAT 1535 1539 5.
 FT DOMAIN 1547 1560 7 X 2 AA TANDEM REPEATS OF K-E.
 SQ SEQUENCE 1585 AA; 180421 MM; 4689B620D2484961 CRC64;

Query Match 5.7%; Score 121.5; DB 1; Length 1585;
 Best Local Similarity 21.0%; Pred. No. 2.7;
 Matches 71; Conservative 50; Mismatches 114; Indels 103; Gaps 17;

QY 51 PPPPP-----NNNNNNSKHTGHSACVPMTERRDELSSEINLRKYMKOSEE 102
 DB 333 PSSSPPTOSDIFENNNNNNNNNNNNN-----NNNNNNNNNNNNNNNEELINNNN 387
 QY 103 NNNLOSQYKITEENTTLREGV-----TPEDEDDIELNG-----AAA 142
 DB 388 NNN--DENKYTEETESLELEKEKLENEEREKILKEKNEIDNLKKNNHLSKGYMRAC 445
 QY 143 AAAPPPIEECPEDLP--EKFDGNPMLAPFMAOQIFMEKSTRDSVDVRYCFYTS 199
 DB 446 NASNDGLEE--EDIPQDEHETNVLTLPCRHNVKVPSSSS--SIDSI----- 492
 QY 200 MATGAARKASAKLESHYLMHNPAPFMEM--KVFEDPQREVAKKIRLRQMGQSV 257
 DB 493 ---ROLWASQKMOGHNLKEKEFFTLRMCNKDVVFDOD-----TPLGHL 535

QY 258 IDYSNMFQMTAQLDWNNEPA-----LIDQYHESLSPHIOE-----LSHEVAKSLS 304
 Db 536 IQY-----NLNNPNPQKPTNIKLELY-----LEDELCKERLVYDLSQSLINNGRP 580
 QY 305 AL-----IGCIIHERRRLAARAAARPRS-PPRALVLP 337
 Db 581 SIWKSHTIDVLSFNKRLREIAMLAKPOSNVPAARLP 618

RESULT 11

Y182_HUMAN STANDARD; PRT; 1157 AA.

AC Q14687;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein KIAA0182 (Fragment).
 GN KIAA0182.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Bone marrow;
 RA MEDLINE=96281124; PubMed=8724849;
 RA Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;
 RT "Prediction of the coding sequences of unidentified human genes. V.
 RT The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
 RT analysis of cDNA clones from human cell line KG-1.";
 RL DNA Res. 3:17-24(1996).

CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; D80004; BA11499.1; -.
 KW Hypothetical protein.
 FT NON_TER 1
 FT DOMAIN 591 596 POLY-PRO.
 FT DOMAIN 685 688 POLY-ARG.
 FT DOMAIN 1042 1047 POLY-GLU.
 SQ SEQUENCE 1157 AA; 130323 MW; BDDADA56FA5FB CRC64;

Query Match 5.6%; Score 120.5; DB 1; Length 1157;

Best Local Similarity 21.9%; Pred. No. 2.1;

Matches 95; Conservative 52; Mismatches 124; Indels 163; Gaps 25;

QY 23 LHPHRSRTAGRSPTPTV-----LGPDCPP-----PPPP-----PNNNNNN 63
 Db 354 LHGRGATATERGRPSDLPPTRAEKLKDAISLAPKQPHLVPVPHHTPPLSLISNG 413
 QY 64 SKHTGHSACVPNNTERRDLSEINNLREKVKOSPENNNLSOVOKLEENTTLREQ 123
 Db 414 IFSLPSSAATALLIORTNE---EKKIARORRLROKEEDR--OSQVSEFQO--VLEQH 466
 QY 124 VE-----PTPEDEDDDELGAATAAAPPPIEECPEDLPKFDGNDMLAP-----FMAQ 175
 Db 467 LDMGRPVPAEAENRPE-----STTRPGRNHREGGRPPRGHGPRLPLSPKQLHNAF 521
 QY 176 COLFMESTRDSVDYRVVCTVTSMATGRARMAKLENSHYTMHTNPAC----- 226
 Db 522 TALNNPVSIMD-----NTLETRRA-----ESHSL-HSHPAFEPSRQAAV 560
 QY 227 -MMMKHVFEDPORREYAKKIRRLRGMSGVIVYSNAFQIADLDWNEPALIDQY--- 282
 Db 561 PLVAVERYF-CPEKAEGPRK-----REPAPLDKTOPP 592

QY 283 -----HESLSPHIOELSH-----LEVAKSLSALIGCIIHERRRLAARAAAR--- 324
 Db 593 PPPREGS-----LEHOFPLDPCGFPAELERSQFTILGQ-----QASLDPQATPCE 641
 QY 325 -----KPRSP-----PRATVLPILASHHGYDPT---EPVGAEMRLTQE---EKERRKL 368
 Db 642 LSGPLKGSFYRRPVPA-----PDPAYTIDEFLQGRRLVSKLDLERRRREA 690

QY 369 NLCLYCCTGGHYAD 382

Db 691 Q-----EKGYVD 698

RESULT 12

AF6_HUMAN STANDARD; PRT; 1816 AA.

ID AF6_HUMAN
 AC P55196; 075087; 075088; 075089; 09WU92;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE AF-6 protein.
 GN MLT4 OR AF6.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]

RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=94061833; PubMed=8242616;
 RA Prasad R., Gu Y., Alder H., Nakamura T., Canaan O., Saito H.,
 RA Huebner K., Gale R.P., Nowell P.C., Kuriyama K., Miyazaki Y.,
 RA Croce C.M., Canaan E.;
 RT "Cloning of the ALF-1 fusion partner, the AF-6 gene, involved in
 RT acute myeloid leukemias with the t(6;11) chromosome translocation.";
 RL Cancer Res. 53:5624-5628(1993).

CC CC [2]
 CC SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.

RC TISSUE=Fetal brain;
 RX MEDLINE=98344142; PubMed=9679199;
 RA Saito S., Matsushima M., Shirahama S., Minaguchi T., Kanamori Y.,
 RA Minami M., Nakamura Y.;

RT "Complete genomic structure, DNA polymorphisms, and alternative
 RT splicing of the human AF-6 gene.";
 RL DNA Res. 5:115-120(1998).

RP SEQUENCE OF 337-1816 FROM N.A. (ISOFORM 2).

RA Williams S.;

RU Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

CC CC -1- FUNCTION: MAY ACT AS AN INTRACELLULAR SIGNALING COMPONENT
 CC CONTROLLED BY RAS SIGNALING PATHWAYS.

CC CC -1- SUBUNIT: BINDS DIRECTLY TO ZO-1 AND OCCUDIN.

CC CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).

CC CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS: 1, 2 (SHOWN HERE) AND 3; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.

CC CC -1- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY A CHROMOSOMAL
 CC TRANSLOCATION T(6;11)(Q27;Q23) THAT INVOLVES MLT4 AND MLL/HRX.

CC CC THE RESULT IS A ROGUE ACTIVATOR PROTEIN.

CC CC -1- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.

CC CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.

CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AB011399; BAA32484.1; -.
 DR EMBL; AB011399; BAA32483.1; -.
 DR EMBL; AB011399; BAA32485.1; -.
 DR EMBL; U02478; AAC50059.1; -.


```

DR EMBL: AF049698; CAB76850.1; -
DR HSSP: 012923; 3PDZ.
DR MIM: 139539; -.
DR InterPro: IPR002710; DIL.
DR InterPro: IPR000253; FHA.domain.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR000159; RA.
DR Pfam: PF01843; DIL; 1.
DR Pfam: PF00498; FHA; 1.
DR Pfam: PF00595; PDZ; 1.
DR Pfam: PF00788; RA; 2.
DR ProDom: PD003376; DIL; 1.
DR SMART: SM00240; FHA; 1.
DR SMART: SM00228; PDZ; 1.
DR SMART: SM00314; RA; 2.
DR PROSITE: PS50106; PDZ; 1.
DR Chromosomal translocation;
KW Chromosomal translocation;
FT DOMAIN 36 206
FT DOMAIN 425 491
FT DOMAIN 804 910
FT DOMAIN 991 1077
FT DOMAIN 162 174
FT DOMAIN 1349 1356
FT DOMAIN 1371 1376
FT DOMAIN 1561 1571
FT SITE 26 26
FT VARSPLIC 1588 1611
FT VARSPLIC 1612 1816
FT VARSPLIC 1666 1743
FT VARSPLIC 1744 1816
FT CONFLICT 373 373
FT CONFLICT 391 391
FT CONFLICT 744 744
FT CONFLICT 1031 1031
FT CONFLICT 1408 1408
SQ SEQUENCE 1816 AA; 205604 MW; EBIFF7F04879CEBF CRC64;

Query Match 5.68; Score 120.5; DB 1; Length 1816;
Best Local Similarity 19.78; Pred. No. 3.6;
Matches 70; Conservative 46; Mismatches 159; Indels 81; Gaps 11;

31 TAGRSPPTPTVTLGP---DCPPPPPPPPNNNNN-----NNSKHTGHSACV 74
1328 TPAATATATPAVASOPTRTDLP PPPPPPPPVYAGDFDGMKMDLPLPPPPPSANQIGLPSAOV 1387
75 PNNTERRRDELSEINNLREKVKQSENNNLQSOVKLTLENTTLREQVEPPEDEDD 134
1388 AAERKRREHQWYE--KKKARLEERERKRREQERKLGQMT---QSLNPAFPFLTA 1442
135 IELRGAAAAAAPP---IEECPEDLPEKFDGCPMDLAPFMAQOCITMEKSTRDSVDR 191
1443 QQCKPKPPTLQRPQETVIRELQPPQCPRIE-----RRDQYITVSK 1485
192 VRCFCPTSMVTGAARMAWASKLERSHYLMNRYAFMMEMKHVFEDDPQRRAVAKRRLR 251
1486 EELSSDSDSPDPWKDKAKETLKQOO-MIIVMLSKIELOLQSKPDPSAESDRLEKLM 1544
252 QGMSYIDYSNAFQMIADLDNMEPALDQYHGLSDHIOEELSHLEVAKSALIGOCI 311
1545 -----LEMOFQKRLQSKQKDEDEDEDDVD-----TMLNQRL 1580
312 HIERRLAARAAARKPPSPRALVLPHTASHNOVPTPEVGGARRMLTOEKKERRR 367
1581 EAERRARLQDEERR-----RQOOLEEMKRREAD-----RAQOEERRRQOE 1621

```

```

RESULT 13
CHDM_DROME STANDARD; PRT: 1982 AA.
AC 097159; Q9YWS50;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Chromodomain helicase-DNA-binding protein M1-2 homolog (dml-2).
GN M1-2 OR CG8103.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN 11
RP SEQUENCE FROM N.A., FUNCTION, AND MUTAGENESIS OF GLY-737.
RX MEDLINE=9055400; PubMed=9836641;
RA Kehle J., Beuchle D., Treuhelt S., Christen B., Kennison J.A.,
RA Bienz M., Muller J.;
RT "dml-2, a hunchback-interacting protein that functions in Polycomb
RL Science 282:1897-1900(1998).
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnikler S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.-H.C., Blazer J.R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrial J.F., Agbayani A., An H.-J., Andrews-Plannkott C., Baldwin D.,
RA Ballew R.M., Sasu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegyan C.,
RA Jajalal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.W., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
CC -1- FUNCTION: VITAL ROLE IN DEVELOPMENT. PROTEIN BINDS TO A PORTION OF
CC HUNCHBACK (HB) PROTEIN THAT IS CRITICAL FOR REPRESSION OF BITHORAX
CC COMPLEX (BXC) GENES. MAY ALSO FUNCTION IN POLYCOMB GROUP (PCG)
CC REPRESSION OF HOX GENES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC -1- SIMILARITY: CONTAINS 2 'CHROMO' DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.

```


CC COAT MUCILAGE.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN DEVELOPING TRICHOMES.
 CC -1- SIMILARITY: BELONGS TO THE HD-ZIP FAMILY OF HOMEBOX PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L32873; AAC80260.1; -.
 DR EMBL: Z54356; CAA91183.1; ALT_INIT.
 DR EMBL: AC011717; AAG52245.1; -.
 DR HSSP: P22808; INK3.
 DR TRANSFAC: T02961; -.
 DR InterPro: IPR001356; Homeobox.
 DR InterPro: IPR002913; START.
 DR Pfam: PF00046; homeobox.1.
 DR Pfam: PF01852; START.1.
 DR SMART: SM00389; HOX.1.
 DR SMART: SM00234; START.1.
 DR PROSITE: PS00027; HOMEBOX.1; 1.
 DR PROSITE: PS00071; HOMEBOX.2; 1.
 DR Homeobox; DNA-binding; Nuclear protein.
 FT DOMAIN 39 60 ASP/GLU-RICH (ACIDIC).
 FT DOMAIN 70 87 ASP/GLU-RICH (ACIDIC).
 FT DNA_BIND 99 158 HOMEBOX.
 SQ SEQUENCE 745 AA; 82988 MM; EAA2DPAAC155D5F3 CRC64;

Query Match 5.6%; Score 120; DB 1; Length 745;
 Best Local Similarity 18.1%; Pred. No. 1.4; Indels 106; Gaps 17;
 Matches 67; Conservative 64; Mismatches 133;

QY 57 NNNNNNSKTHGKSCAC-----VPMTERRDESEI-----NNLR 93
 DB 92 NKGTAKRRKKYHHTTQIRIMEALFKETHPDEKQKQSLKGLAPROYKFFQNR 151
 QY 94 E-KVMKSENNNLSQVOKLTENTTLRQVE-----PTPEDEDDIELRGA----- 140
 DB 152 TQIKAIQRHNSILKALEKLEENKAMRESFSKANSQRCGGGPDHLNENKLA 211
 QY 141 -----AAAAAPRPTEECRDELPKFGNPDMLAFMAQCOIEMEKSTRDSEVDVRVC 195
 DB 212 LDKLFAALGRTTPYPLQASCSDDQEHRL-GSLDFY-----TGVFALEKSRIA 256
 QY 196 FVTSMTGRRARMAA-----KLRSHYLMHNPAPFMEMKHYFEDPQREVAKRIR 248
 DB 257 EISNATLELOKMATSGPMLRVSVEGREIL-NYDEYLKE--FPOAQASSPGRRTI 311
 QY 249 RLROGMSGVIDYSNAFQIAADLD--WNE-----PALIDQYHEG-----LSDHTQ- 291
 DB 312 EASRNAGIV--FMDAHKLAQSFMDVCGMKERFACILSKAATVDVLRQGEGRIDGATQL 369
 QY 292 -----BELSHLEVANSLS--ALIGQCIHIE-RLIARAANAARPPSPRA 332
 DB 370 MFGEMQLTTPVVPTRFVYFVSCROLSPKMAIVDVSVSEDSNTEKESALLKCRKLP 429
 QY 333 LVLPRIASHH 342
 DB 430 CIIEDTSMGH 439

RESULT 15
 TAGB_DICDI STANDARD; PRT: 1905 AA.
 AC P54683; 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE prestalk-specific protein tagB precursor (EC 3.4.21.-).
 GN TAGB.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=AX4;
 RX MEDLINE=95262903; PubMed=7744252;
 RA Shaulsky G., Kuspa A., Loomis W.F.;
 RT "A multigene resistance transporter/serine protease gene is required
 RT for prestalk specialization in Dictyostelium.";
 RL Genes Dev. 9:1111-1122(1995).
 CC -1- FUNCTION: INTERCELLULAR COMMUNICATION VIA TAGB MAY MEDIATE
 CC INTEGRATION OF CELLULAR DIFFERENTIATION WITH MORPHOGENESIS.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO PEPTIDASE FAMILY
 CC S8; ALSO KNOWN AS THE SUBTILASE FAMILY.
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ATP-BINDING
 CC TRANSPORT PROTEIN FAMILY (ABC TRANSPORTERS). MDR SUBFAMILY.
 CC -1- SIMILARITY: STRONG, TO TAGC.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U20432; AAA62212.1; -.
 DR HSSP: P13569; INBD.
 DR DictyDb: DD02059; tagB.
 DR InterPro: IPR003593; AAA.
 DR InterPro: IPR001140; ABC_transporter_tmam.
 DR InterPro: IPR003439; ABC_transporter.
 DR InterPro: IPR001687; ATP_GTP_A.
 DR InterPro: IPR000209; Peptidase_S8.
 DR Pfam: PF00064; ABC_membrane.1.
 DR Pfam: PF00005; ABC_tran.1.
 DR PRINTS: PRO0723; SUBTILISIN.
 DR SMART: SM00362; AAA.1.
 DR PROSITE: PS00136; SUBTILASE_ASP; FALSE_NEG.
 DR PROSITE: PS00137; SUBTILASE_HIS.1.
 DR PROSITE: PS00138; SUBTILASE_SER.1.
 DR PROSITE: PS00211; ABC_TRANSPORTER.1.
 DR HydroLase; Serine protease; ATP-binding; Transport; Transmembrane;
 KW Signal.
 FT SIGNAL 1 31
 FT CHAIN 32 1905 PRESTALK-SPECIFIC PROTEIN TAGB.
 FT DOMAIN 378 700 PROPEASE.
 FT DOMAIN ? 1905 ABC_TRANSPORTER.
 FT TRANSMEM 1011 1031 POTENTIAL.
 FT TRANSMEM 1076 1096 POTENTIAL.
 FT TRANSMEM 1121 1141 POTENTIAL.
 FT TRANSMEM 1210 1230 POTENTIAL.
 FT TRANSMEM 1309 1329 POTENTIAL.
 FT TRANSMEM 1332 1352 POTENTIAL.
 FT ACT_SITE 387 387 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 432 432 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 695 695 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1553 1560 ATP (POTENTIAL).
 FT DOMAIN 63 67 POLY-GLN.
 FT DOMAIN 95 104 POLY-ASN.
 FT DOMAIN 107 134 POLY-ASN.
 FT DOMAIN 311 321 POLY-SER.
 FT DOMAIN 833 837 POLY-SER.
 FT DOMAIN 838 844 POLY-GLY.
 FT DOMAIN 871 876 POLY-LEU.
 FT DOMAIN 1012 1015 POLY-ILE.
 FT DOMAIN 1386 1389 POLY-GLU.
 FT DOMAIN 1398 1404 POLY-GLY.
 FT DOMAIN 1445 1450 POLY-ASN.
 FT DOMAIN 1765 1779 POLY-ASN.


```

QY 120 LRSOVPTEDEDDDETLRGAAGAAAPPPPIEECEPDLPEKFDGNPDLAPFAOQIF 179
    |||
DB 61 LRSOVPTEDEDDDETLRGAAGAAAPPPPIEECEPDLPEKFDGNPDLAPFAOQIF 120
QY 180 MEKSTRDFSVDRVAVCEVTSMMTGRAARWASAKLERSHYLMHNPAPFMEMKHVEDPQR 239
    |||
DB 121 MEKSTRDFSVDRVAVCEVTSMMTGRAARWASAKLERSHYLMHNPAPFMEMKHVEDPQR 180
QY 240 REVAKRIRRLRQMGSVIDYNSAFQMIADLDMNEPALIDQYHEGLSDHIOEELSHLEV 299
    |||
DB 181 REVAKRIRRLRQMGSVIDYNSAFQMIADLDMNEPALIDQYHEGLSDHIOEELSHLEV 240
QY 300 AKLSALIGOCIHIERLARAARARPRSPRALVLPFHASHHQVDPTEPVGARMRLTQ 359
    |||
DB 241 AKLSALIGOCIHIERLARAARARPRSPRALVLPFHASHHQVDPTEPVGARMRLTQ 300
QY 360 EKERERRKMLCYCGTGHYADNCPAKAKSSPAGNSPAPL 401
    |||
DB 301 EKERERRKMLCYCGTGHYADNCPAKAKSSPAGNSPAPL 342

```

```

RESULT 2
Q96A68 PRELIMINARY; PRT; 325 AA.
ID 096A68
AC 096A68;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, last sequence update)
DE ME3-Like 1 (PATERNALLY EXPRESSED GENE 10 ORF1).
GN ME3L1 OR PEG10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Nakamura Y.; Furukawa Y.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Satoh S.; Furukawa Y.;
RL "Isolation of ME3 like gene 1.";
RN [3]
RP Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=21218929; PubMed=11318613;
RA Ono R., Kobayashi S., Wagatsuma H., Aisaka K., Kohda T.,
RA Kaneo-Ishino T., Ishino F.;
RT "A Retrotransposon-derived Gene, PEG10, Is a Novel Imprinted Gene
RT Located on Human Chromosome 7q21.";
RL Genomics 73:232-237(2001).
DR EMBL; AB049150; BAB68387.1; -.
DR EMBL; AB049834; BAB43951.1; -.
SQ SEQUENCE 325 AA; 36965 MW; 118EACFA97F2A76 CRC64;

```

Query Match 79.6%; Score 1703; DB 4; Length 325;
 Best Local Similarity 100.0%; Pred. No. 3e-123;
 Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 77 MTERRRDELSEINLRKVKVAKOSEENNLOSOKLFEENTTLREQVEPTPEDEDDIE 136
    |||
DB 1 MTERRRDELSEINLRKVKVAKOSEENNLOSOKLFEENTTLREQVEPTPEDEDDIE 60
QY 137 LRGAAGAAAPPPPIEECEPDLPEKFDGNPDLAPFAOQIFPEKSTRDSVDRVAVCE 196
    |||
DB 61 LRGAAGAAAPPPPIEECEPDLPEKFDGNPDLAPFAOQIFPEKSTRDSVDRVAVCE 120
QY 197 VTSMTGRAARWASAKLERSHYLMHNPAPFMEMKHVEDPQREVAKRIRRLRQMG 256
    |||

```

```

DB 121 VTSMTGRAARWASAKLERSHYLMHNPAPFMEMKHVEDPQREVAKRIRRLRQMG 180
QY 257 VIDYNSAFQMIADLDMNEPALIDQYHEGLSDHIOEELSHLEVAKLSALIGOCIHIER 316
    |||
DB 181 VIDYNSAFQMIADLDMNEPALIDQYHEGLSDHIOEELSHLEVAKLSALIGOCIHIER 240
QY 317 LARAAARPRSPRALVLPFHASHHQVDPTEPVGARMRLTQEEKERRKMLCYCGT 376
    |||
DB 241 LARAAARPRSPRALVLPFHASHHQVDPTEPVGARMRLTQEEKERRKMLCYCGT 300
QY 377 GGHYADNCPAKAKSSPAGNSPAPL 401
    |||
DB 301 GGHYADNCPAKAKSSPAGNSPAPL 325

```

```

RESULT 3
Q9EQ11 PRELIMINARY; PRT; 231 AA.
ID Q9EQ11
AC Q9EQ11;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, last sequence update)
DE MYELIN EXPRESSION FACTOR-3-LIKE PROTEIN (FRAGMENT).
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC SRRAIN-C57BL/6;
RX MEDLINE=21105984; PubMed=11158386;
RA Volft J.-N., Koertling C., Scharf M.;
RT "My3/Gypsy retrotransposon fossils in mammalian genomes: did they
RT evolve into new cellular functions?";
RL Mol. Biol. Evol. 18:266-270(2001).
DR EMBL; AF302691; AAC39979.1; -.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00098; zf_CCHC; 1.
DR SMART; SM00343; Znf_C2HC; 1.
KW zinc-finger.
FT NON_TER 1
SQ SEQUENCE 231 AA; 26171 MW; DEAB2A2E624F3974 CRC64;

```

Query Match 44.1%; Score 943; DB 11; Length 231;
 Best Local Similarity 77.0%; Pred. No. 5.9e-65;
 Matches 181; Conservative 20; Mismatches 30; Indels 4; Gaps 2;

```

QY 167 DMLAPFAOQIFPEKSTRDFSVDRVAVCEVTSMMTGRAARWASAKLERSHYLMHNPAP 226
    |||
DB 1 DMLAPFAOQIFPEKSTRDFSVDRVAVCEVTSMMTGRAARWASAKLERSHYLMHNPAP 60
QY 227 MMEKHVFEDPQREVAKRIRRLRQMGSVIDYNSAFQMIADLDMNEPALIDQYHEGL 286
    |||
DB 61 MMEKHVFEDPQREVAKRIRRLRQMGSVIDYNSAFQMIADLDMNEPALIDQYHEGL 120
QY 287 SDHIOEELSHLEVAKLSALIGOCIHIERLARAARARPRSPRALVLPFHASHHQVDP 346
    |||
DB 121 NPDIRAEISRQEAFTLALITACIHIERLARAARARPRSPRALVLPFHASHHQVDP 176
QY 347 TEPVGARMRLTQEEKERRKMLCYCGTGHYADNCPAKAKSSPAGNSPAPL 401
    |||
DB 177 TEPVGARMRLTQEEKERRKMLCYCGTGHYADNCPAKAKSSPAGNSPAPL 231

```

```

RESULT 4
Q98SV9 PRELIMINARY; PRT; 704 AA.
ID Q98SV9
AC Q98SV9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, last annotation update)
DE GAG-PROTEASE.

```

OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 OC Tetraodontidae; Takifugu.
 OX NCBI_Taxid=31033;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-SUSHI-SAN RETROTRANSPOSON;
 RX MEDLINE-21321259; PubMed-11428463;
 RA Butler M., Goodwin T., Simpson M., Singh M., Poulter R.;
 RT "Vertebrate LTR retrotransposons of the Tf1/sushi group.";
 RL J. Mol. Evol. 52:260-274(2001).
 DR EMBL: AF316578; AAC60684.1; -
 DR MEROPS: A02.UPW.; -
 DR InterPro: IPR001969; Asp.-protease.
 DR InterPro: IPR001995; Asp.-prot.-retrov.
 DR InterPro: IPR000953; Chromo.
 DR Pfam: PF00098; zf-CCHC; 1.
 DR SMART: SM00298; CHROMO; 1.
 DR SMART: SM00343; ZnF_C2HC; 1.
 DR PROSITE: PS00141; ASP_PROTEASE; UNKNOWN_1.
 DR PROSITE: PS50175; ASP_PROT_RETROV; 1.
 DR PROSITE: PS50013; CHROMO_2; 1.
 DR Protease; Zinc-finger.
 KW SEQUENCE 704 AA; 76619 MW; 81E2ELC9AB0153B3 CRC64;
 SQ
 Query Match 17.4%; Score 373; DB 13; Length 704;
 Best Local Similarity 31.6%; Pred. No. 1.7e-20;
 Matches 95; Conservative 51; Mismatches 127; Indels 28; Gaps 6;
 OY 105 NLGSGVQKLEENTTL-----REQVEPTPEDEDDIELRGAATAAPPIEECEPDLP 159
 DB 29 DIMASLQSLNTSVYDLSLIGWAEQVSPPAPEALEVYQRYAAA-----PMEHPHP--1P 81
 OY 160 EKFGNPNMLAPFMAOQOIFMEKSTRDPSVDVRVRCFTSMATGGRARMAAKLEKSHYL 219
 DB 82 ERYSGEACVCSFLQCLGVFDLPITYPGDRAKAFAYVNLISGRAAWATVALENOPPA 141
 OY 220 MHNTPAFMEMKHYFEDQREVRARIRLRGSGSVIDYSNAFQMTADLDNNEPALI 279
 DB 142 SSSPEFAAEKRVFDDHPDQSGEAAASQIRSLRGSSSVADY---FRILAAKSGMNDTLR 198
 OY 280 DQYHEGLSDHQEELSHLEVAKSLALIGOCIHIERLRARAAAARPPSPRALVLPPIA 339
 DB 199 GVFQGLAEALKDELATREESGDLTLLSLAIQIDNRLRERHROSNRGMPRESLSVSGR 258
 OY 340 SHHGYD-----PTEPVGARM-----RLTOEKERRRKLNLCTGCTGGHYADNCPA 386
 DB 259 SGAYVESSTLAPAPPOEDVDDEPMQLGPRSLSHAERIRRSANCLCNGQAHFVADCPY 318
 OY 387 K 387
 DB 319 Q 319
 RESULT 5
 OY3283 PRELIMINARY; PRT; 371 AA.
 AC OY3283;
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE GAG POLYPROTEIN.
 OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 OC Tetraodontidae; Takifugu.
 OX NCBI_Taxid=31033;

RN [1]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-SUSHI-ICHI RETROTRANSPOSON;
 RX MEDLINE-98382517; PubMed-9714821;
 RA Poulter R., Butler M.;
 RT "A retrotransposon family from the pufferfish (fugu) Fugu rubripes.";
 RL Gene 215:241-249(1998).
 DR EMBL: AF030881; AAC33525.1; -
 DR InterPro: IPR001878; ZnF_CCHC;
 DR Pfam: PF00098; zf-CCHC; 1.
 DR PRINTS: PR00939; C2HCZNFINGER.
 DR SMART: SM00343; ZnF_C2HC; 1.
 KW Polyprotein; Zinc-finger.
 SQ SEQUENCE 371 AA; 40540 MW; 617926FC010730AC CRC64;
 Query Match 17.0%; Score 364.5; DB 13; Length 371;
 Best Local Similarity 27.6%; Pred. No. 3.4e-20;
 Matches 101; Conservative 62; Mismatches 144; Indels 59; Gaps 8;
 OY 68 GHKSACVNMTERRDDELSEELNLRKVMKQ-----SENNNLSQVQ-----K 112
 DB 12 GPRTPPLSPLERRVREASQSLSSQSELTKAFPTIQEISELQSSQSTTSLSALNSQ 71
 OY 113 LTEENTTLREQVEPTPEDEDDIELRGAATAAPPIEECEPDLP--PEKFDGNPDMLA 170
 DB 72 MSAMATVLASTITQKIGSP-----GGAAPSEPLSPRAEPRLAPRPFGGFDLCK 124
 OY 171 PMAOQOIFMEKSTRDPSVDVRVRCFTSMATGGRARMAAKLEKSHYLMHNPAMMEM 230
 DB 125 GFLHQCELLFHPQPSRFVSDAKYGFITSLADKLSWAIAVADDPRLSSDYSAFREF 184
 OY 221 KHVEPDQRRFARAKIRLRGSGSVIDYSNAFQMTADLDNNEPALIDQYHEGLSDHI 290
 DB 185 KAVEFHPYGEDAASRLALQOQSRSAVEYDFEFLIAESHWGCTALRSAYRRLSEAI 244
 OY 291 QEELSHLEVAKSLALIGOCIHIERLRARAAAARPP----- 327
 DB 245 -KDLVRPSPSLNLTLSLQMDRLRERQRAQAGSGTRQLSHRTSSAPDFSLST 303
 OY 328 -SPRALVLPPIASHHGYDPT-----EPVGGARMRLTOEKERRRKLNLCTGCTGGHYA 381
 DB 304 AAFP-----PHLLQSPAPHPSPVGEPMQIGRSLRSQREORLRLDCLCYGNNGHFI 358
 OY 382 DNCPAK 387
 DB 359 QACPVR 364
 RESULT 6
 OY6V43 PRELIMINARY; PRT; 1252 AA.
 AC OY6V43;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE POLYPROTEIN (FRAGMENT).
 OS Aspergillus flavus.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 OX NCBI_Taxid=5059;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NRRL 19997; TRANSPOSON-AFRTL-1;
 RX MEDLINE-95314240; PubMed-7793909;
 RA McAlpin C.E., Mannarelli B.;
 RT "Construction and characterization of a DNA probe for distinguishing strains of Aspergillus flavus.";

RL Appl. Environ. Microbiol. 61:1068-1072(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 19997; TRANSPOSON=AFRTL-1;
RA Okubara P.A., Tibbot B., McAlpin C.E., Hua S.-S.T.;
RT "Afrt-1, a retrotransposon-like element in the aflatoxin-producing
fungus Aspergillus flavus.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF362957; AAL26311.1; -
FT NON_FER 1252 1252
SQ SEQUENCE 1252 AA; 144752 MW; DERDCSDDD86C4E8 CRC64;

Query Match 13.0%; Score 279; DB 3; Length 1252;
Best Local Similarity 24.3%; Pred. No. 5,7e-13;
Matches 94; Conservative 72; Mismatches 155; Indels 66; Gaps 14;

QY 58 NNNNNKHHGHKS-----ACVPNTERRRDELSEINNLREKVKQSEENNLL 106
DB 2 :SSQSSSKKPPVKSTPPAETDSESETTKKQOLKQMSKITQVLNNNAKE-----NQEIENL 57
QY 107 QSQVQKLTLEENTTLREQVEPTPEDEDDIELRGAAAAAPRP-EEECPEDLPEKEDGN 165
DB 58 KVV-----LGEAEIRINEQODHIAQDAVGSAPKDALGKVKLP--AEPDGT 105
QY 166 PDLAPPAQOLIFMEKSTDFSVDRVVCFTSMGTGRARMAASAKLERSHY--LMHNY 223
DB 106 RSKQAQALTLQNNMIHANRKLIDEADKVFIFSTHLGAAANNFEPYI-REYEVVPDNW 164
QY 224 PAFPMEM-----KIV---FEDPQREYAKKIRLRLOGMSGVIDYNAFQMIADLD 272
DB 165 SNTTRELTFDSGLRKHLERTFFGDVDAEVAVERKLQDY--RGSASTYAAEFQOIIISMD 223
QY 273 WNEPALIDYHEGLSDHIOELSHLEVAKSLALIGOCIHIERLARA-----AAARK 325
DB 224 WNEGLCVNLHOMSGHVKDEFARIDRAPATILEAIDRAVKVNNRYHERLMKKRNEARK 283
QY 326 PRSPRALVLT-----PHIASHQVDPTPEVGGARMLTQEEKERRRKLNTLC 371
DB 284 GSHRPKQGYKSNDRERTGVKNDPYGPKPMELDATEGQSK-GISQKERERRRREKLC 342
QY 372 LYCGTGGHYADNCPAKSKSPAGNSP 398
DB 343 YNCGRAGHMSKDCRQKRN-SOPANRKP 368

RESULT 7
ID 000833 PRELIMINARY; PRT; 853 AA.
AC 000833;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE GAG POLYPROTEIN.
GN GAG.
OS Fusarium oxysporum.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; mitosporic Hypocreales; Fusarium.
OX NCBI_TaxID=5507;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F. SP. LYCOPERSICI 42-87;
RA MEDLINE=96132549; PubMed=8544829;
RA Anaya N., Roncero M.I.;
RT "skippy, a retrotransposon from the fungal plant pathogen Fusarium
oxysporum.";
RL Mol. Gen. Genet. 249:637-647(1995).
DR EMBL: L34658; AAA88790.1; -
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam: PF00098; zfc-CCHC; 1.
DR PRINTS: PR00939; C2HCZNFINGER.
DR SMART: SM00343; Znf_C2HC; 1.
KW Polypeptide; Zinc-finger.

SQ SEQUENCE 853 AA; 99422 MW; 1AEDD18CFBBA5B06 CRC64;

Query Match 11.2%; Score 240.5; DB 3; Length 853;
Best Local Similarity 21.2%; Pred. No. 3.3e-10;
Matches 85; Conservative 65; Mismatches 157; Indels 93; Gaps 14;

QY 33 GRSPPTVTYLGPPCPPPPPPNNNNNNKATGKSACVPMTERRDEL---SEET 89
DB 19 GHQPAAP-----ANPAVVRPRPTDQNMDDADDSQSSD--DSEVRLREQLGNTNEM 69
QY 90 NNLREKV-----MKQSEENNIQSOVOKITEENTLRL--EQVEPTPEDEDDIELRG 139
DB 70 NEMRQMLEEFTALOHQONQSNNTQOEMYNLASANNRDPGEVLKPS----- 118
QY 140 AAAAAAPRPIEECPDLEPEKFGNDMLAPFAAOQIFMEKSTDFSVDRVVCVTS 199
DB 119 -----PEYFDGTPSKLPTFLQSRAPFTTYIPNOFRNDSAKVMYAG 159
QY 200 MMTGRARMAASAKLERSHYLMHNY---PAFPM-----MKHVPEDPQ 238
DB 160 RLITTAQWQRP-----IMNDYMTNPPYKIQRTALLFGENGHMEALKAFTID 212
QY 239 RREYAKKIRLRLOGMSGVIDYNAFQMIADLWNEPALIDYHEGLSDHIOELSHLE 298
DB 213 EKGQAEKRIKTLKQ--TGSASTLGEFFQLASKLPWDODVLMSEFFDALKEVOOELWEKD 271
QY 299 VAKSLSLIGOCIHIER---LARAARPRSPRALVLRPHIASHQVDPT--EPVGG 352
DB 272 RPRTLVEYINNAVATIDRQAPWRTNRSNGKGRQDNKPRYHANAGRTROTDTSTGTAGP 331
QY 353 ARMRLTQEEKERRRKLNTLCYCGTGGHYADNC--PAKASK 390
DB 332 MAMGTRKRSK-----VTCYNCGKKGHYEBECKNPVATNQ 367

RESULT 8
ID 09HFY8 PRELIMINARY; PRT; 837 AA.
AC 09HFY8;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE GAG PROTEIN.
GN GAG.
OS Colletotrichum gloeosporioides (Anthracoze fungus) (Glomerella
cingulata).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetes; Incertae sedis; Phyllachorales; Phyllachoraceae;
OX NCBI_TaxID=5457;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON=RETROTRANSPOSON CGRET;
RA Zhu P., Oudemans P.V.;
RT "A long terminal repeat retrotransposon Cgret from the phytopathogenic
fungus Colletotrichum gloeosporioides on cranberry.";
RL Curr. Genet. 0:0-0(2000).
DR EMBL: AF264028; AAG24791.1; -
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam: PF00098; zfc-CCHC; 1.
DR SMART: SM00343; Znf_C2HC; 1.
KW Zinc-finger.

SQ SEQUENCE 837 AA; 97738 MW; EFLD4BC70FD55003 CRC64;

Query Match 10.7%; Score 229.5; DB 3; Length 837;
Best Local Similarity 20.6%; Pred. No. 2.2e-09;
Matches 76; Conservative 68; Mismatches 174; Indels 51; Gaps 8;

QY 68 GHKSACVP-NMTERRRDELSEINNLRL---EKYMKQSEENNNIQQSQVQKLTLENTT--- 119
DB 13 GSSSROYPNMLDGLQKMKDDSDSEDETEELKQLEKTNNDPKEMKDFNONAALAKE 72


```

          ID      Q94T04       PRELIMINARY:             PRT:    1862 AA.
AC         O94T04;
DT        01-DEC-2001 (TREMBLrel, 19, Created)
DT        01-DEC-2001 (TREMBLrel, 19, Last sequence update)
DE        01-DEC-2001 (TREMBLrel, 19, Last annotation update)
DR        PUTATIVE RETROELEMENT.
OS        Oryza sativa (Rice).
OC        Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC        Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC        Ehrhartioideae; Oryzeae; Oryza.
NCBI_TaxID=4530;
RP        SEQUENCE FROM N.A.
RC        STRAIN-NIPPONBARE;
RT        McCormie W.R.;
RL        "Rice genomic sequence.";
RN        Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RM        [1]
RP        SEQUENCE FROM N.A.
RC        STRAIN-NIPPONBARE;
RT        McCormie W.R.;
RL        Submitted (MAR-2001) to the EMBL/Genbank/DDBJ databases.
RM        [1]
RN        RN
RP        SEQUENCE FROM N.A.
RC        STRAIN-NIPPONBARE;
RT        Palmer L.E., Bal H., Cordum H., Johnson D., Minx P., de la Bastide M.,
RA        Nascento L.U., Spiegel L.A., Kirchhoff K.A., King L., Preston R.R.,
RA        Vill M.D., Baker J.P., Miller B., Santos L., Zutavern T., Kult K.H.,
RA        Rodriguez S., Cunniss D.M., Ballia V.S., Shah R.S., Bahret A.,
RA        O'Shaughnessy A., Dedhia N.N., McCormie W.R.;
RL        Submitted (MAY-2001) to the EMBL/Genbank/DDBJ databases.
RM        [4]
RN        RN
RP        SEQUENCE FROM N.A.
RC        STRAIN-NIPPONBARE;
RT        Palmer L.E., Bal H., Cordum H., Johnson D., Minx P., de la Bastide M.,
RA        Nascento L.U., Spiegel L.A., Kirchhoff K.A., King L., Preston R.R.,
RA        Vill M.D., Baker J.P., Miller B., Santos L., Zutavern T., Kult K.H.,
RA        Rodriguez S., Cunniss D.M., Ballia V.S., Shah R.S., Bahret A.,
RA        O'Shaughnessy A., Dedhia N.N., McCormie W.R.;
RL        Submitted (AUG-2001) to the EMBL/Genbank/DDBJ databases.
DR        EMBL, AC022352; AAA51585.1; -.
SQ        SEQUENCE   1862 AA;  211396 MW;  291B106D83539AB4 CRC64;

Query Match               8.6%; Score 185; DB 10: Length 1862;
Best Local Similarity     20.2%; Pred. No. 1.6e+05;
Matches 97; Conservative 56; Mismatches 133; Indels 194; Gaps 22
OY      5 RVLTAKRRSRGGGDDPGILPHRSSEATAGSRSEPPTVTGLDPCCPPPPPNNNNNNS 64
            |:|::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      356 RWVTTNRANNTGDGNQPECSNHNG--NPpp-----pppppppppDTN----- 398
OY      65 KATHGSACVPMTTERRDDELSEEINNLREKYWKOSEENNLIQSOUOKLTEENTTLREQV 124
            | : | :: | :: | :: | :: | :: | :: | :: | :: | :: | :: | :: |
Db      399 -----ATTTOI-----LAQAANNM-----TAFLHLL 419
OY      125 EETPEDDEDIDLGAASAAAAPPPIEECECPDL---PEKF--DGMP-DMLAPYA---- 174
           | : | :: | :: | :: | :: | :: | :: | :: | :: | :: | :: | :: |
Db      420 QAMPPOON-----APPPOHSKLAELRIPTPTESSSNPNVDALMDLHAVGKK 467
OY      175 ---OCOIFEMEKSTRDEFSDRVACEFTVSMMTGAARW-----ASAK 212
           |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      468 LGTYVCS-----DEEKVIFAHOLGPSPLMWDHPFOAQOPPEGOPTMAWFTA 515
OY      213 IERSHYLMHNYPAFPMMEKNHFVEDQRREVARKRKIRLRLOGMGSVTDYSNAQMIA---- 268
           | : | :: | :: | :: | :: | :: | :: | :: | :: | :: | :: | :: |
Db      516 FERTH---VVAGVAALK-----KRFERELKGGRSNAVWEYLHEFNLNARVAP 558

```

QY	269	OLDMNEPALIIOYHEGLSDHIOELSHLEVA---KSLSLAGGCHTE-----R	315
Db	559	EDVRDE-----EKEKFLAGMDPELSVRLISGDYPPDFQLADKSLRLAKHKELESKR	613
QY	316	RA-----RAAARKRSPRALVYLPHIASSHQVD-----	345
Db	614	RLANFRNOGAGNRVRYTHPRPGYSSSQOQOQOQOQPRSAFPRGFVYVRYPQDQOQOQST	673
QY	346	-----PTEPVGAGRMRLIOEERERRKMLCLYCGTGHYADNCP---AKASKSSPAGNS	397
Db	674	RAPRPPTPVGCGGRDRDAGQOOR-----LCFNCEFGPHFADKCEKPRRQOQADAPRSNN	728
RESULT 14			
Q9AYCO		PRELIMINARY;	PRT; 1781 AA.
AC	Q9AYCO;		
DT	01-JUN-2001 (TREMBLrel. 17, Created)		
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	POLYPROTEIN.		
GN	OSJNBA0094H10.15.		
OS	Oryza sativa (Rice).		
OC	Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;		
OC	Ehrhartoideae; Oryzeae; Oryza.		
XX	NCBI_TaxID=4530;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=NIPONBARE;		
RA	Spiegel L.A., Nascimento L.U., de la Bastide M., Kirchoff K.A.,		
RA	King L., Preston R.R., Vil M.D., Baker J.P., Miller B., Zutavern T.,		
RA	Rodriguez S., Santos L., Kuit K.H., Cummins D.M., Bailja V.S.,		
RA	Shah R.S., Bairet A., Bal H.P., O'Shaughnessy A., Dedhia N.N.,		
RA	Mcombe M.R.;		
RT	"Genomic Sequence For Oryza sativa, Niponbare Strain, Chromosome X,		
RT	Clone OSJNBA0038E19, complete sequence."		
RL	Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.		
DR	EMBL; AC080019; AAK13118.1; "		
DR	InterPro: IPR001969; Asp_protease.		
DR	InterPro: IPR001584; Rve.		
DR	InterPro: IPR000477; RVTse.		
DR	InterPro: IPR01878; Znf_CCHC.		
DR	Pfam; PF00665; Ive; 1.		
DR	Pfam; PF00078; rvt; 1.		
DR	Pfam; PF00098; zf_CCHC; 1.		
DR	SMART; SM00343; znf_C2HC; 1.		
DR	PROSITE; PS00141; ASP_PROTEASE; UNKNOWN.1.		
KW	RNA-directed DNA polymerase; Zinc-finger.		
SO	SEQUENCE 1781 AA; 202845 MW; 9984A707A125FE8A CRC64;		
Query Match 8.6%; Score 184.5; DB 10; Length 1781;			
Best Local Similarity 20.1%; Pred. No. 1.6e-05;			
Matches 95; Conservative 56; Mismatches 141; Indels 181; Gaps 19;			
QY	5	RVLTAKRRRSRGSGDDPGLHPRSEATAGRSPPTVTLGPPCPPPPPPNNNNNS	64
Db	285	RKVMTRRNAAITSDBGQNPEDSNHNNHG---SEPP-----PPPPPPPDITN-----	327
QY	65	KHTGKHSACVPMPTERRRDELSEELINNLREKVMKQSEENNINLQSQVQKLTENTTLREOV	124
Db	328	-----AIIQLILAQOANMMTAPLHLL-----QNPQGHN-----	355
QY	125	EPTPEDEDDDLRLCAAAAAAAPPRIEECCPDL---PEKF---DGNP-DMLAPMA-----	174
Db	356	-----APPPPOHSLKIAELRICPTPTSSNNPNVDALDMLHAVGKK	396
QY	175	-----OCQIFMEKSTRDFSVDRVRCVFTSMKMGRAARW-----ASAK	212
Db	397	LDTVQCS-----DEKVIFFAAHQLOGPRLSLMMDHQAOTQPSQPTTWARFTIA	444
QY	213	LERSHYLMHNTPAFMEKMKHVEDPQREVARKRITRLROGMSVIDYSNAFQMTA----	268

```
Db 445 FRFTH-----VPAGVALK-----KREFRELKOGNRSVMEYLHEFNMLATAP 487
QY 269 QDDMNEPALIDYHGLSHIOELSHLEVA---KSLSALIGOCIHIE-----R 315
Db 488 EDREDE-----EKQKFLAGMDELVSCLSGDYPDFORLVDSIKLEKKHKELESKKR 542
QY 316 RLA-----RAAARPPRPPRALVLPRIASHHOVDPTPEV 350
Db 543 RLNFHNGOGANRVRYTNPYGGSSSQOQQOQPPRSAPRPFQVYVRPQPOEQNOQGT 602
QY 351 GGAR-----MLTQEKERKRRKLNLCLYCGTGHYADNCPAKASKSSPAGNSP 398
Db 603 RARPPRPPTVQPGGRRDAOGPQLCFNCFEPGHFADKCPKPRROGOQAPPRP 655
```

```
RESULT 15
Q94H22 PRELIMINARY; PRT; 1473 AA.
AC Q94H22;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PUTATIVE POLYPROTEIN.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Buell C.R., Yuan Q., Ouyang S., Moffat K.S., Hill J.N., Gansberger K.,
RA Brenner M., Burgess S., Hance M., Shvartsbeyn M., Tsilfin T.,
RA Riggs F., Hsiao J., Zismann V., Blunt S., Pai G., Vanaken S.E.,
RA Uteback T.R., Feldblyum T.V., Quackenbush J., Salzberg S.L.,
RA White O., Fraser C.M.;
RT "Oryza sativa chromosome 3 BAC OSJNB4007G22 genomic sequence.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC084831; AAK52169.1; -.
KW Polypeptidein.
SQ SEQUENCE 1473 AA; 169336 MW; 9978B784571B22B4 CRC64;
```

Query Match 8.5%; Score 182.5; DB 10; Length 1473;
Best Local Similarity 21.6%; Pred. No. 1.9e-05;
Matches 88; Conservative 59; Mismatches 125; Indels 135; Gaps 19;

```
QY 55 PPNNNNNNSKHTGHSACVPMTERRDELSEIINLREKVMKQSENNMLQSOVKLT 114
Db 265 PNNNNNNNGEN-----PTLAQVLAQ-AQLMMNMQOLOQONQGNHAPPONKLV 315
QY 115 EENTLREQVPEPTDEDDDELRLGAAAAAARPPRIEECEPELPEKFGNPDMLAPMA 174
Db 316 E-----FLRVPRPTFSSITTNPVEAGDWLHALEKLD-----LL 348
QY 175 QCOIFMEKSTRDFSVDRVRCFVTSMGTGAAR-----ASAKLERS 216
Db 349 QC-----TDOEKVSFAHQHGPASEWMDHFLNRTTAEPITWLEFTAAPRKT 396
QY 217 HYLMMNTPAFMEMKHYFEDPQREVAKKRTRLRQGMGSYIDYSNAPOMIA-----QDLD 272
Db 397 H-----IPSGVVSILK-----KKEFRSLTQGSRSVTEYLHEFNRLARYAPEDVR 439
QY 273 WNEPALIDYHGLSDHIOELSHLEVA---KSLSALIGOCIHIE-----RRLAR 319
Db 440 TDEQRO-ERFLEGLND-----ELSTPLMTG DYHDFQKLVDAKIQEDKYNRMEOKKRRIAH 494
QY 320 AAAARKPPRPRALV-----LPHIAHHQVDP-----TEPVGGARMRLT 358
Db 495 FKAQOGNSQRPRLTLGQSMPOGSSSVYRQRFNNMAGNNIRNOQAPRYAAS-----T 550
QY 359 QEE---KERRRKLNLCLYCGTGHYADNCP-AKASKSSPA-GNSPAP 400
```

```
Db 551 QQQPARKEQGIKPGVCFNCGDPGHYADKCPKPRRVVVPQSNSTAP 597
Search completed: October 4, 2002, 15:40:49
Job time: 151 sec
```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2002, 15:38:18 ; Search time 14.69 Seconds
(without alignments)
666,757 Million cell updates/sec

Title: US-09-631-863A-2
Perfect score: 2139
Sequence: 1 MRKKRVLTAKTKRRSGRGOD.....DNCPAKAKSSSPAGNSPAPL 401

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Issued_Patents_AA:*
2: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/6C.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfilest1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	159	7.4	146	US-09-209-525-2	Sequence 2, Appli
2	155	7.2	126	US-09-209-525-45	Sequence 45, Appli
3	149.5	7.0	110	US-09-209-525-52	Sequence 52, Appli
4	125	5.8	905	US-08-574-959A-9	Sequence 9, Appli
5	125	5.8	905	US-09-357-014-9	Sequence 9, Appli
6	125	5.8	1135	US-08-574-959A-7	Sequence 7, Appli
7	125	5.8	1135	US-09-357-014-7	Sequence 7, Appli
8	120	5.6	80	US-09-209-525-54	Sequence 54, Appli
9	118.5	5.5	2293	US-09-368-590-2	Sequence 2, Appli
10	118	5.5	834	US-08-539-205A-6	Sequence 6, Appli
11	115	5.4	1651	US-09-540-245A-18	Sequence 18, Appli
12	114	5.3	640	US-09-026-343-2	Sequence 2, Appli
13	112	5.2	1612	US-08-545-860D-48	Sequence 48, Appli
14	112	5.2	1612	PCT-US94-04496-48	Sequence 48, Appli
15	110.5	5.2	538	US-09-370-368-9	Sequence 9, Appli
16	109.5	5.1	1939	US-09-310-187A-1	Sequence 1, Appli
17	107	5.0	418	US-09-026-587-1	Sequence 1, Appli
18	107	5.0	418	US-09-227-420-1	Sequence 1, Appli
19	107	5.0	1829	US-09-157-420-1	Sequence 1, Appli
20	106.5	5.0	1099	US-09-442-100-2	Sequence 2, Appli
21	106	5.0	1872	US-08-188-582-14	Sequence 14, Appli
22	106	5.0	1872	US-08-646-715-14	Sequence 14, Appli
23	106	5.0	1886	US-08-938-105-3	Sequence 3, Appli
24	105.5	4.9	657	US-09-370-368-7	Sequence 3, Appli
25	104.5	4.9	393	US-09-026-587-3	Sequence 3, Appli
26	104.5	4.9	393	US-09-227-420-3	Sequence 3, Appli
27	104	4.9	2414	US-08-227-536-2	Sequence 2, Appli

28	104	4.9	2414	5	PCT-US95-04682-2	Sequence 2, Appli
29	104	4.9	3969	4	US-08-061-376-5	Sequence 5, Appli
30	102.5	4.8	413	3	US-09-089-879-7	Sequence 7, Appli
31	102.5	4.8	1137	3	US-08-369-043-2	Sequence 2, Appli
32	102.5	4.8	1719	2	US-08-459-568-4	Sequence 4, Appli
33	102.5	4.8	1719	2	US-08-399-411-4	Sequence 4, Appli
34	102.5	4.8	1719	3	US-08-516-859A-4	Sequence 4, Appli
35	102	4.8	2843	1	US-07-741-940-2	Sequence 2, Appli
36	102	4.8	2843	1	US-08-289-548A-2	Sequence 2, Appli
37	102	4.8	2843	1	US-08-452-654-2	Sequence 2, Appli
38	102	4.8	2843	2	US-08-370-235A-2	Sequence 2, Appli
39	101.5	4.7	1103	3	US-09-162-373-1	Sequence 1, Appli
40	101.5	4.7	1103	4	US-09-467-946-1	Sequence 1, Appli
41	101	4.7	1162	2	US-08-728-323A-2	Sequence 2, Appli
42	100.5	4.7	1388	4	US-09-572-191-2	Sequence 2, Appli
43	100.5	4.7	2842	1	US-07-741-940-7	Sequence 7, Appli
44	100.5	4.7	2842	1	US-08-289-548A-7	Sequence 7, Appli
45	100.5	4.7	2842	1	US-08-452-654-7	Sequence 7, Appli

ALIGNMENTS

```

RESULT 1
US-09-209-525-2
; Sequence 2, Application US/09209525
; Patent No. 6303770
; GENERAL INFORMATION:
; APPLICANT: Lok, Si
; APPLICANT: Konklin, Darrell C.
; TITLE OF INVENTION: Mammalian Alpha Helical Protein-1
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/209,525
; CURRENT FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-209-525-2

Query Match      7.4%; Score 159; DB 4; Length 146;
Best Local Similarity 27.4%; Pred. No. 3.3e-07;
Matches 45; Conservative 21; Mismatches 68; Indels 30; Gaps 2;

QY 83 DELSEINNLREKVKMKOSENNNLSOVOKLPEENTTLREOVETPEDEDDDIETRGAA 142
   |||::||:| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 3 DELVLLHLLMKRRALSIENSQLEQLRLVLCERASLLROYRP----- 46
QY 143 AAAPPPPIEECPEDLPKPFQGNPDMLAFMAQCOIFMEKSTRDFSVDRVCFTSMNT 202
   || | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 47 -----PSCPVPPEPFENGSRLPEFIVQTASYMLVNRRCNDAMKVAFLISLTF 97
QY 203 GRAARMAAKLERSHYLMHNYPAFMEKHYF-----EDPQRR 241
   | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 98 GRAEEMVVPYIEMDSPILGDYRAFLDEMKQCGWDDDDDDDE 141

RESULT 2
US-09-209-525-45
; Sequence 45, Application US/09209525
; Patent No. 6303770
; GENERAL INFORMATION:
; APPLICANT: Lok, Si
; APPLICANT: Konklin, Darrell C.
; APPLICANT: Parrish, Julia E.
; TITLE OF INVENTION: Mammalian Alpha Helical Protein-1
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/209,525
; CURRENT FILING DATE: 1998-12-10

```



```

;      TYPE: amino acid
;
;      TOPOLOGY: linear
;
;      MOLECULE TYPE: protein
;
;      SEQUENCE DESCRIPTION: SEQ ID NO: 7:
;
US-09-357-014-7

```

Query Match	5.88;	Score 125;	DB 4;	Length 1135;
Best Local Similarity	20.58;	Pred. NO. 0.0075;		
Matches 45;	Conservative 22;	Mismatches 75;	Indels 78;	Gaps 7;

```

Qy      22 GLHPRKSEATAGRSPTPTVTIGP-----DCPPPPPPPP----- 56
      ||| :||| || |||||
Db      790 GLPPLPPPSGATPP-PIAPYGPPTASPPVPAKEEPEELPAAPGPLPPPPPPPPYPGP 848

```

```

QY 57 -----NNNNNNNSKHTGHKACVPMNTERRDE-- 84
      ::: : | : |
Db 849 VXLPPQLVPEGTPGGGCPPALEEDLTVININSNDEEEEGEEEEE 908

```

```

QY      85  -----LSEELNNLRKVKMGQSENNNNNQSQVKLTTEENTTLRGVEPTPEDEDDLE-- 136
          || : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      909  EEEEDFEEEDDEEYFEEEEEEEEEEEF--EEGCLSEEEEEDEDEEEEFEEV 965

```

```

QY 137 -----LRGAAAAAAPPPPIEEEC--PELLEKRFNDGND 167
      |         |||||         | : | : | :
Db 966 EDLEFGTAGGEVEEGAPPPPTLPPLALPPPESSPKVQPEPE 1005

```

RESULT 8
US-09-209-525-54
Sequence 54. Application ITS/09200525

```

; Patent NO. 6303770
; GENERAL INFORMATION:
; APPLICANT: LOK, Si
; APPLICANT: Confidin Darrell C

```

```

: APPLICANT: PATISH, JULIA E.
: TITLE OF INVENTION: Mammalian Alpha Helical Protein-1
: FILE REFERENCE: 97-71
: CURRENT APPLICATION NUMBER: US/09/209,525

```

```

; CURRENT FILING DATE: 1998-12-10
;
; NUMBER OF SEQ ID NOS: 56
;
; SOFTWARE: FastSeq for Windows Version 3.0
;
; SEQ ID NO 54

```

```

; LENGTH: 80
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-209-525-54

```

Query Match	5.68;	Score 120;	DB 4;	Length 80;
Best Local Similarity	32.98;	Pred. No. 0.00053;		

	conservative	mismatches	indels	gaps
159 PEKFDGNPDMALPFMAQCQIFMEKSTRDPSVDRYVCFTYSMTGTGARRNASKLESHY	23	14	37	0

00	2	FELTNOSSKULU EL YUGOSLAVIYANANT SUDAPALVAT DISTILLIGNEBMY FILLIMUSF	01
QY	219	LMIHNYPAFMEMKHVE	234
:	:	:	:

RESULT 9

Sequence 2, Application US/09368590
Patent No. 6187563
GENERAL INFORMATION:

APPLICANT: SOLIMENA, MICHAEL
TITLE OF INVENTION: INTERACTING POLYPEPTIDES FOR
TITLE OF INVENTION: AUTOANTIGENS OF AUTOIMMUNE DISEASES
FILE REFERENCE: 101918-200 (OCR-941)

; CURRENT AFFILIATION NUMBER: 03/03/2008, 250
 ; CURRENT FILING DATE: 1999-08-04

```

; EARLIER APPLICATION NUMBER: 60/095,657
; EARLIER FILING DATE: 1998-08-07
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2293
; TYPE: PRT
; ORGANISM: Human
US-09-368-590-2

```

Query Match	5.5%;	Score 118.5;	DB 4;	Length 2293;
Best Local Similarity	23.3%;	Pred. No. 0.079;		
Matches	91;	Conservative	41;	Mismatches 145;
			Indels	113;
			Gaps	20;

```

QY      29  EATGRSPPPTV-----TLGPCPPPPPPPPNNNNNNNSKHTGHKSACVNMTE 79
      : | : ||| : | | | : | : | :
Db    1846  KAEQSKQPPPLTGRKFGDPTLAAKAPLLRP-----GTYERGLEP-LAR 189

```

```

QY      80 RRRLDSEELNNLRKVMKQSENNNNQSQVQLTEENTLTREQVEP-----TPEDEDD 134
      |  | | | :  :: :  | | :: : | |  | | | :
Db     1892 RASPTLSAEVRT-RGVYVRQELKPERLOPRIDRLPE----IPGRVEPALPAAPEDAET 194

```

```

QY      135 TELGAAAAAAPPPTIEEC--PEDUEKFDGNPMDLAPFMAQOQIFMESTRDSVDRV 192
          ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db      1947 PATPAAAEQVFRPRERQESADRAEELPR-----RRPERQESVDOS 198

```

```

QY 193 RCFVTSMGGAARASAKJERSHYTMHNPYAFMEMKHFVEDORREVAKRJLRQ 252
      : | | | | | : | | | : | | |
Db 1988 E-----EARRRRPERQSAHEHAHSLTLGR-----YEOMERRR--ERRERLER 203

```

```

QY 253 GMSYIDISARQWLAQDLDWNEPALDQIHGSDSDHIOEELSHUEVAASLSALIOQJIN 312
      : | | | : | | | : | | | :
Db 2032 Q-----ESSEQEMPITGDLVYKCATLAD-----IVEQLQENAGPGLPA--GPSLP 207

```

```

QY      313  LERKLAARAAARKPKSPKRALVLFHSHNOVDTEFGGAKRMFLVLELEKRRKLNLC 372
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2076 QPREL-----PFGRLPNGLELPERTR-----PDRP--RARDNPKRRRRPRPRE----- 211

```

```

07 3/3 1CGGNNIADCFAANASASSFA--GNSAF 400
      | ||      :: ||:|| |::|||
Db 2118 -GGEG-----GSRRSRAPAGGSAPAP 2140

```

RESULT 10
US-08-539-205A-6
: Sequence 6, Application US/08539205A

```

; patent NO. 9001015
;
; GENERAL INFORMATION:
;
; APPLICANT: Beach, David H.
;
; APPLICANT: Callajuri, Maureen
;

```

APPLICANT: Netisky, Bradley
TITLE OF INVENTION: Ubiquitin Ligases, and Uses Related Thereto
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:

ADDRESS: FOLEY, HOW & ELLIOTT
STREET: One Post Office Square
CITY: Boston
STATE: MA

COUNTRY: USA
ZIP: 02109-2170
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```

;
; COMECON: IBM PC compatible
;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```

AFFILIATION NUMBER: 03/06/2005
 FILING DATE: 04-OCT-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Vincent, Matthew P.

REGISTRATION NUMBER: 30/05
; REFERENCE/DOCKET NUMBER: CSV-005.01

TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 832-1000
TELEFAX: (617) 832-7000
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 834 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-539-205A-6

Query Match 5.5%; Score 118; DB 3; Length 834;
Best Local Similarity 22.2%; Pred. No. 0.021;
Matches 50; Conservative 29; Mismatches 66; Indels 80; Gaps 12;

OY 50 PPPPPP-----NNNNNNSKHTGHSACVPMTERRDELSSEINLRE 94
DB 70 PPPPLPPGMEKVDNLGRTYYVNHNNRTTQWHR-----PSL-----MDVSESDNNITRQ 118
OY 95 KVMQSENNNLOQVOKLRENTTLRQVETPEDED-----DIELRG-AAAA 143
DB 119 --INQEAHRRFRSR-----RHISEDLEPESEGVDPEPWETISEEVNIAGDSLGL 168
OY 144 AAPPPPIE-----EECPEDLPEKFDGNPMLAPFMAQCOIFMEKSTRDSVDRVYCF 196
DB 169 ALPPPVSPGSRSTSQELSEELSRRLQITPD-----SNGEFSLLIRREPS-SRLRSCS 221
OY 197 VTSMT-----GRAARWASAK--LERSHYLMHN 222
DB 222 VTDAVAEQHLPSPVAVYHTTTPGLPSCGMEERKDAKGRTYYVNNH 266

RESULT 11
US-09-540-245A-18
Sequence 18, Application US/09540245A
Patent No. 6270984
GENERAL INFORMATION:
APPLICANT: Goodman, Corey
APPLICANT: Kild, Thomas
APPLICANT: Brose, Katja
APPLICANT: Tessier-Lavigne, Marc
TITLE OF INVENTION: Modulating Robo: Ligand Interactions
FILE REFERENCE: B98-031-3
CURRENT APPLICATION NUMBER: US/09/540, 245A
CURRENT FILING DATE: 2000-03-31
PRIOR APPLICATION NUMBER: 60/065, 544
PRIOR FILING DATE: 1997-11-14
PRIOR APPLICATION NUMBER: 60/081, 057
PRIOR FILING DATE: 1998-04-07
NUMBER OF SEQ ID NOS: 20
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 18
LENGTH: 1651
TYPE: PRT
ORGANISM: human
US-09-540-245A-18

Query Match 5.4%; Score 115; DB 4; Length 1651;
Best Local Similarity 19.0%; Pred. No. 0.1;
Matches 81; Conservative 47; Mismatches 171; Indels 128; Gaps 15;

OY 14 SGRGGDGLPHRSEATAGRSPTTYTLGDCPPPPPPNNNN-----NNNSKHTGH 69
DB 1153 SDRGSSTSGSGHKKKATPPVKOGGMWADLLPPPAHPPHNSSEYINVSDESYDQ 1212
OY 70 KSAC--VNMTERRDELSSEINN-----LREKVMQSENNNLOQVOKLRENTTLR 121
DB 1213 EMPGVPAPARMYLDQDELESEEDERGPTTPVKGAASSPAVSYSHQSTATLTPSPQELQ 1272
OY 122 EOVETPEDE--DDLELRGAAAAAAP-----PIEECPEDLPEKFDG 164

DB 1273 PMLQDPEETGHMQHQRRRQPVSPPPPPRISPPTHYGISGPLVSDMDTDAREEED 1332
OY 165 NPDMLAPFMAQCOIFME--KSTRDSVDRVRYCFYTSMTTGAARWASAKLERSHYLMHN 222
DB 1333 EADMEVAKQTRRLRLRLEIQTBPASSVGDLESSYTGSMING--WGSASEDN----- 1382
OY 223 YPAFMENKHFVEDPQREVAKRIRRLROGMSYIDVSNAFQMIADLDNNEPALIDQY 282
DB 1383 -----ISSGRSSVSSDGSF--FTPADFAQAVAAAAE 1412
OY 283 HEGLSDHIOELSHLEVAKSLSALIGOCITERRLARAAAAKRPSPRALVLPPIASHH 342
DB 1413 YAGLK-----VARQMODAAGR-----HFHASQ 1436
OY 343 QVDPEPV-----GAGARLTOEKEKRRKLNLCITGCTGCH-----YADNCPKAKAS 391
DB 1437 CRRPTSPVSTDSNMSAAVQKTRPAKKLKHQ-----PQHLRETYTDLDP-PPVP 1486
OY 392 SPAGNSP 398
DB 1487 PPAIRSP 1493

RESULT 12
US-09-026-343-2
Sequence 2, Application US/09026343
Patent No. 6008018
GENERAL INFORMATION:
APPLICANT: DUAN, D. ROXANNE
APPLICANT: SHILATIPAR, ALI
APPLICANT: CONAWAY, JOAN W.
APPLICANT: CONAWAY, RONALD C.
TITLE OF INVENTION: ELI2, A New Member of an ELI Family of
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESS: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/026, 343
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/038, 447
FILING DATE: 19-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: GOLDSTEIN, JORGE A.
REGISTRATION NUMBER: 29,021
REFERENCE/DOCKET NUMBER: 1488, 0880001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 372-2540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 640 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-026-343-2

Query Match 5.3%; Score 114; DB 3; Length 640;
Best Local Similarity 19.4%; Pred. No. 0.034;
Matches 64; Conservative 55; Mismatches 125; Indels 86; Gaps 13;

[illegible]

RESULT 13
 US-08-545-8600-48
 Sequence 48: Application US/08545860D
 Patent No. 6040140
 GENERAL INFORMATION:
 APPLICANT: Croce, Carlo
 APPLICANT: Canaan, Eli
 TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
 TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
 TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 Region
 NUMBER OF SEQUENCES: 94
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
 ADDRESSEE: No. 6040140rls
 STREET: One Liberty Place, 46th floor
 CITY: Philadelphia
 STATE: Pennsylvania
 COUNTRY: USA
 ZIP: 19103
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/545,860D
 FILING DATE: 07-MAR-1996
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/US94/04496
 FILING DATE: 22-APR-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/US92/10930
 FILING DATE: 09-DEC-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/327,392
 FILING DATE: 19-OCT-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/320,559
 FILING DATE: 11-OCT-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/062,443
 FILING DATE: 14-MAY-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/971,094
 FILING DATE: 30-OCT-1992
 PRIOR APPLICATION DATA:

```

1 APPLICATION NUMBER: US 07/888,893
2 FILING DATE: 27-MAY-1992
3
4 PRIOR APPLICATION DATA:
5 APPLICATION NUMBER: US 07/805,093
6 FILING DATE: 11-DEC-1991
7 ATTORNEY/AGENT INFORMATION:
8 NAME: Deluca Esq., Mark
9 REGISTRATION NUMBER: 33,229
10 REFERENCE/DOCKET NUMBER: TJU-1262
11 TELECOMMUNICATION INFORMATION:
12 TELEPHONE: (215) 568-3100
13 TELEFAX: (215) 568-3439
14 INFORMATION FOR SEQ ID NO: 48:
15 SEQUENCE CHARACTERISTICS:
16 LENGTH: 1612 amino acids
17 TYPE: amino acid
18 STRANDEDNESS: single
19 TOPOLOGY: linear
20 MOLECULE TYPE: protein
21 HYPOTHEetical: NO
22 ANTI-SENSE: NO
23
24 US-08-545-860D-48

```

Query Match	5.28;	Score 112;	DB 3;	Length 1612;
Best Local Similarity	19.78;	Pred. No. 0.19;		
Matches	60;	Conservative 41;	Mismatches 138;	Indels 66;
				Gaps 9;

[illegible]

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/04496
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Deluca Esq., Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: TJU-1242
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 1612 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
PCT-US94-04496-48
```

```
Query Match 5.2%; Score 112; DB 5; Length 1612;
Best Local Similarity 19.7%; Pred. No. 0.19;
Matches 60; Conservative 41; Mismatches 138; Indels 66; Gaps 9;
```

```
QY 31 TAGSSPTPTVTLGP---DQPPPPPPPPNNNN-----NNSKHTGKSKCY 74
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1329 TPAIIPTPAVASQPIITDLPDPPPPPHYAGFDGKSMDLPLPPPSAQCIGLPSQV 1388
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 75 PNMTERRDELSEIINNIREKVMKQSENNNLOSQVOKLTEENTTLREQVPTPEDEDD 134
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1389 AARARRRREHQRWE--KERAPLEERERKRREGKRLGMR---QSLNPAFSPPLTA 1443
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 135 TELRGAAAAAAPP---IEECPEDLPEKFDGNPDLAPMACQIFMEKSTDFSYDR 191
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1444 QQMPKPSSTLQRPQETVIRLQPOQOPRYE-----RRDLQYITVSK 1486
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 192 VRVCFTSMKMGRAARMAAKLESHYLMNYPAFMMEMKHVFEDPORREYAKKRIRLR 251
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1487 EELSSGSLSPDPKRAKKELEKQQ-MHLYDMLSKETIQLOSKPPDRSAESPRKLKM 1545
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 252 QGMGSVIDYSNAFQMIADLDWNEPALIDYHEGLSDHIQELSHLEVAKSLALIGQCI 311
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1546 -----LEMOFQRLQESKQKDEDEDEDEDDVD-----TWLIMQRL 1581
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 312 HIERR 316
| | |
DB 1582 EAERR 1586
```

```
RESULT 15
US-09-370-368-9
; Sequence 9, Application US/09370368
; Patent No. 6258932
; GENERAL INFORMATION:
; APPLICANT: Anders Valine
; TITLE OF INVENTION: PEPTIDES THAT BLOCK VIRAL INFECTIVITY
; TITLE OF INVENTION: AND METHODS OF USE THEREOF
; FILE REFERENCE: TRIPEP-003A
; CURRENT APPLICATION NUMBER: US/09/370,368
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Moloney Murine Leukemia Virus
US-09-370-368-9
```

```
Query Match 5.2%; Score 110.5; DB 4; Length 538;
Best Local Similarity 19.0%; Pred. No. 0.057;
Matches 96; Conservative 51; Mismatches 184; Indels 175; Gaps 22;
```

```
QY 21 PGLPHRS-----EATGRSPTPTVTLGPCPPPPPP-----PPNNNN 61
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 71 PGPGHDPDQVPIYVTEALADPPP---WVPPVHPKAPPPLLPASPLPLEPLS--- 123
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 62 NNSKHTGKSACVPMNTERDELSEIINNIREKVMKQSENNNLOSQVOKLTEENTTLR 121
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 124 -----TPQSSLYPALNPSL-----GAKPKRQVLSDSGGL---IDLTEDPPIYR 166
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 122 EOVEPTPEDEDDI-ELRGAAAAAP-----PPIEECPEDLPEKFDGNPDW 168
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 167 DP-RPPSPDRDSDGEATPAGEADPPSPMASRLKGRREPVAADSTTSQAFPLRFGNGQL 225
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 169 -LAPFMAQCQIFMEKSTDFSVDRVY-CFYTSMTGTGAARW----- 208
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 226 QYWPFSDDLVMKNNNPFSESDPGKLTALIESVLTHTQPTMDCQQLLTLLGEEKOR 285
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 209 -----ASAKLER-----SHYLMH----- 221
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 286 VLEARKAVRGDDGRPTQLPNEVDAAFLERPDMEYTTQAGRNLVHTRQLLIGLONAG 345
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 222 NYPAFMMEMKHVFEDPORREYAKKRIRLRQMGSVIDS---NAFQMIADLDWNEPA 277
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 346 RSPINLAKVKIGITQGP--NESPSAFLERLKAVRYRTPYDEDPGQETVNSMFTWQAP 403
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 278 LIIDYHEGLSDHIQELSHLEVAKSLALIGQCHI-----ER 315
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 404 DIGKLERL-----EDLRN---KTLGDLVREARIEFKRETPEEREIRIREBEKEER 454
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 316 RLARAAAARKRSPRALVLPRIASHHOVDTEPVGARMRLTOEKKRRK---LNLCL 372
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 455 RTDEDEQKEKRRDNR-----HREMSRLATVYVSGQNDQEGRRRSQDDCCT 505
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 373 YCGTGHTADNCPAKASKSPAGNSP 398
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 506 YCEBQGHMAKDCPRR--PRGPRGPRP 529
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: October 4, 2002, 15:38:43
Job time: 25 sec

